Search for Planckian Black Holes in the Di-Lepton Channel with the ATLAS Detector at the LHC

Silje Raddum

THESIS SUBMITTED FOR THE DEGREE
MASTER OF SCIENCE

Department of Physics
University of Oslo

June 2011

Abstract

In some scenarios proposing extra dimensions, the fundamental Planck scale is in the order of a TeV, and the apparent weakness of the gravitational force is a consequence of the large compactified volume of the extra dimensions. These scenarios render possible the non-perturbative process of black hole formation at hadron colliders. It has been argued that black hole signatures based on thermal multi-particle final states are very unlikely. However, strong gravity effects at center of mass energies of the order of the Planck mass are expected to yield an increase in the $2 \rightarrow 2$ production cross section. This thesis reviews the signatures and discovery potential of Planckian black holes, by which is meant true or virtual black holes or simply strong gravity effects, decaying to two leptons in the context of the ADD model in pp collisions at $\sqrt{s} = 7$ TeV at the LHC. Based on data recorded by the ATLAS experiment during 2010 which correspond to a total integrated luminosity of ~ 40 pb^{-1}, no statistically significant excess above the Standard Model expectation is observed. A combined search for high-mass and boosted di-lepton final states results in upper limits at the 95% confidence level on the production cross section for three Planckian black hole models. Assuming six large extra dimensions and a Planck mass of 2 TeV, the quoted limits are; 8.2×10^1 pb for conservation of B, L and flavours; 6.2×10^1 pb for conservation of B and L; 5.3×10^1 pb for conservation of B-L only.

Acknowledgments

First and most of all I would like to thank my supervisor Alex Read for his great guidance, ideas, support and exquisite banana bread for Friday lunches. I am also grateful for the invaluable help provided by my fellow Master students Vanja Morisbak, Magnar Bugge and Kyrre Næss Sjøbæk. Bjørn Hallvard Samset deserves a special thanks for helping me getting started with the GRID and analysis tools used.

Finally, I would like to thank the rest of the EPF group at UiO for answering my many questions and just for being such an extraordinary working place, not just professionally, but also socially. And last, but not least, the EPF group is truly paradise on Earth for cake monsters, so thanks for all the cake!

Introduction

The Large hadron Collider (LHC) is the world's largest and highest-energy particle accelerator. Colliding proton beams, it is expected to address some of the most fundamental questions of physics, and one of the major issues is whether the elusive Higgs boson exists or not. Physicists are also puzzled by the weakness of gravity, its strength being 41 orders of magnitude weaker than the electromagnetic force. This is not necessarily true, and gravity might only *appear* weak due to the existence of large extra dimensions. If that is the case, the LHC might produce microscopic black holes.

In the context of the theory proposed by Nima Arkani-Hamed, Savas Dimopoulos and Gia Dvali in 1998, this thesis will explore the signatures of microscopic black hole production in the Planckian regime with the ATLAS detector at the LHC. The thesis is divided into five chapters and three appendices.

Chapters 1 and 2 are the introductory chapters. In Chapter 1 the Standard Model with its mathematical formalism and shortcomings are briefly introduced. Then follows an introduction to theories of extra dimensions and a theoretical presentation of higher-dimensional black holes. The LHC and the different components of the ATLAS detector are presented in Chapter 2.

In Chapter 3 the principles of event generation, simulation and reconstruction with the ATLAS software are explained. Using BlackMax, a black hole event generator, branching ratios for different decay channels are presented. The relevant background processes of the Standard Model are also discussed.

Using signal samples produced with Atlfast-II, a fast simulation algorithm for ATLAS, a feasibility study is presented in Chapter 4. Here, the baseline cuts from a similar analysis are investigated and applied if found appropriate. It is shown that Planckian black hole production would be visible as an excess of events in the tails of Standard Model distributions. Promising signal-to-background ratios and expected significances are calculated.

An analysis of the data collected by ATLAS in 2010 is given in Chapter 5. Recommended corrections to data and simulations, as well as sources of systematic uncertainties, are presented. Most of the background processes are described by simulations, while others are estimated from data. Finally, using profile likelihood estimation, upper limits at 95% confidence level on the production cross section are derived for three Planckian black hole models assuming six large extra dimensions and a Planck mass of 2 TeV.

Contents

Chapter 1

Theory

1.1 The Standard Model

The most prominent theory of the fundamental particles and their interactions is the Standard Model (SM). It has successfully predicted the outcome of a variety of experiments confirming its tenability. During the 1950s a burst of new particles were discovered due to the development of new particle accelerators. The seemingly chaotic situation led to Wolfgang Pauli's famous remark: "Had I foreseen this, I would have gone into botany." Fortunately the SM breaks this jungle of particles down to a handful of assumed point-like particles describing both matter and forces.

Matter and force-mediating particles are easily distinguished by their spin quantum number S. While matter is made up of $S = 1/2$ *fermions*, forces are mediated by $S = 1$ gauge *bosons*. Each particle possesses an interaction field yielding them the ability to interact with other particles through the exchange of gauge bosons. The fermions are divided into six quarks and six leptons, which may again be grouped into three generations. While the particles of each generation have similar properties, they differ in mass.

The SM contains three fundamental forces, each with its respective force-mediating gauge bosons. *Electromagnetic* interactions occur between electrically charged particles through an exchange of the massless and neutral photon. Similarly the *strong* interactions are mediated by massless, coloured gluons and occur between particles carrying a colour charge, including the gluons themselves. The *weak* force is mediated by the three massive gauge bosons, $W^{\pm}$ and Z^0, and it is the only force in which all fermions, in addition to the massive gauge bosons themselves, interact.

The quarks, which carry colour charge, are the only fermions interacting via the gluon field which is a so-called confining field. An isolated quark has never been observed. They are always confined in compound systems like the proton. Quarks also have electric charge and come in six different flavours: up, down, charm, strange, top and bottom. The up, charm and top quarks are considered as up-type quarks, and similarly the down, strange and bottom quarks are called down-type quarks. The mass eigenstates are not eigenstates of the weak force, and this gives rise to quark mixing between generations through their coupling to the $W^{\pm}$-boson. Experiments have shown that quarks, via the weak force, couple most strongly within the same generation.

The known charged leptons are the electron, the muon and the tau. Each of these are associated with a neutral neutrino. Clarification of the neutrino masses has yet to come, but the observed phenomenon *neutrino oscillation* requires non-zero masses. Apart from this phenomenon, weak interactions between leptons are restricted to leptons within the same generation.

Table 1.1 lists the three fermion generations, while the forces and their respective gauge bosons are listed in Table 1.2.

	First generation	Second generation	Third generation	Charge
Quarks	u (up)	c (charm)	t (top)	$+2/3$
	d (down)	s (strange)	b (bottom)	$-1/3$
Leptons	e (electron)	μ (muon)	τ (tau)	-1
	ν_e	ν_μ	ν_τ	0

Table 1.1: The fundamental fermions of the SM

Force	relative strength	boson	mass	charge	spin
Electromagnetic	1	γ (photon)	0	0	1
Strong	25	g (gluon)	0	0	1
Weak	0.8	$W^\pm$	80.4 GeV	±1	1
		Z^0	91.187 GeV	0	1
Gravitational	10^{-41}	G	0	0	2

Table 1.2: The fundamental forces and their respective gauge bosons of the SM. Note that the graviton G is not (yet) a component of the SM. It is a hypothetical elementary particle predicted to have zero mass, because of the unlimited range of gravitation, and spin 2, because the source of gravity is the stress-energy tensor, a second-rank tensor. The strengths are relative to the electromagnetic force between two up quarks separated by 10^{-18} m.

1.1.1 Theoretical framework

Dirac laid the foundation for *quantum field theory* (QFT) in his famous paper on "The Quantum Theory of the Emission and Absorption of Radiation" [1]. Soon after, in 1928, he formulated the *Dirac equation*,

$$(i\gamma^\mu\partial_\mu - m)\psi(x) = 0, \tag{1.1}$$

which describes $S = 1/2$ material particles, i.e. fermions. By quantization of the electromagnetic field, quantum electrodynamics (QED) was born. Later on, this theory proved to be in amazingly good agreement with experiments. The quantization of the electromagnetic field led naturally to the idea of quantization of other classical fields. This gave rise to a model describing the strong interaction, quantum chromodynamics (QCD) and also the development of a model accounting for both weak and electromagnetic interactions, electroweak theory. The SM is a collection of such quantum field theories describing the interacting fields. What follows is a brief discussion of the basic principles.

The material is mostly drawn from [2], but [3] also proved helpful in summarizing the theory behind the SM.

Symmetry

Symmetry is an essential component of the SM, and its mathematics are provided by group theory. *Noether's theorem*, which states that for every continuous symmetry in nature there is a corresponding conservation law, implies a deep bond between symmetries and physics. A system is said to be invariant under a specific transformation, i.e one or more quantities of the system undergoing change, if the physics of the system remains the same.

The *Lagrangian* L of a dynamical system summarizes the dynamics of the system, whereas Lagrange's equation,

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_j}\right) - \frac{\partial L}{\partial q_j} = 0, \quad j = 1, 2, \ldots, d, \tag{1.2}$$

gives a simple and elegant description of the time evolution of the system. In particle physics, the generalized coordinates q_j are taken to be the particle fields at each point in space. A Lagrangian density $\mathcal{L}$ is introduced, and by shrinking the discrete lattice size $\delta \mathbf{x}_i$ to zero, it is related to the Lagrangian by

$$L(t) = \int d^3\mathbf{x} \mathcal{L}\left(\phi_r, \phi_{r,\alpha}\right). \tag{1.3}$$

The compact notation $\phi_{r,\alpha}$ is defined as

$$\phi_{r,\alpha} = \frac{\partial}{\partial x^\alpha}\phi_r. \tag{1.4}$$

Invariance of the Lagrangian under e.g. translational or rotational transformations lead to conservation of linear or angular momentum, respectively. By postulating a set of symmetries for a given system, the Lagrangian obeying the corresponding conservation laws can be constructed. The SM Lagrangian is constructed by basing this approach on the gauge symmetry group $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$.

QED and conservation of electric charge

The Lagrangian density for a free Dirac field (fermion field),

$$\mathcal{L}_0 = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi, \tag{1.5}$$

is invariant under the *global* phase transformation

$$\psi(x) \rightarrow \psi'(x) = e^{iqw}\psi(x). \tag{1.6}$$

Such transformations are characterized by the transformation parameter ω being independent of the space-time coordinate x. The corresponding conserved quantities are charge q and charge current j^μ. To describe the interactions of (charged) relativistic fermions with the electromagnetic field, one can use the same procedure as in non-relativistic quantum mechanics, namely the substitution

$$\partial_\mu \rightarrow D_\mu = \left[\partial_\mu + \frac{iq}{\hbar c} A_\mu(x)\right], \tag{1.7}$$

with $A_\mu(x) = (\phi, \mathbf{A})$ being the four-vector potential. This substitution is often referred to as the "minimal substitution" and leads to the correct wave equation for a particle of charge q in the electromagnetic field. Applying this to Eq. 1.5 gives

$$\mathcal{L}_1 = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi - q\bar{\psi}\gamma^\mu \psi A_\mu(x), \tag{1.8}$$

where $-q\bar{\psi}\gamma^\mu\psi$ is recognized as the conserved charge current $j^\mu(x)$. From classical electromagnetism it is known that the fields $\mathbf{E}$ and $\mathbf{B}$ remain invariant under the gauge transformation

$$A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu \omega(x), \tag{1.9}$$

where $\omega(x)$ is an arbitrary function. To sustain the invariance of $\mathcal{L}_1$, one must demand that the Dirac field transforms according to the *local* phase transformation,

$$\psi(x) \rightarrow \psi'(x) = e^{iqw(x)}\psi(x). \tag{1.10}$$

By coupling the electromagnetic field to the conserved charge current, the free fermion field has become an interacting field with the property of local phase invariance. The quantum of the gauge field A_μ is the photon, and it can be shown that it is required to be massless for the theory to be gauge invariant.

The group of all phase factors $e^{i\alpha}$, where α is a real parameter, is called U(1). QED is said to be an abelian gauge theory, where the term *abelian* indicates the commutativity of all the elements of U(1). Physically this implies that the coupling is equal for equal charges, regardless of the other properties of the particles involved.

QCD and conservation of colour charge

Studies of hadrons suggested exact colour symmetry and that the force-mediating bosons acting between quarks carry colour charge themselves. This led to the idea that the strong interaction should be described by a non-abelian gauge theory providing conservation of colour charge. In a non-abelian gauge theory there arise self-coupling terms with different coupling strengths in the Lagrangian density. In contrast to the abelian photon field, non-abelian gauge fields carry the charges which the symmetry conserves, hence the self-coupling.

The approach is more or less carried out in the same manner as for QED, but more tedious calculations are needed. What resulted is a theory invariant under $\mathrm{SU}(3)_C$ transformations. QCD contains eight coloured force-mediating gluons and six flavoured quarks represented as colour triplets,

$$\mathbf{q} = \begin{pmatrix} q_r \\ q_g \\ q_b \end{pmatrix}.$$

Electroweak theory and conservation of weak hypercharge and weak isospin

In 1954 Yang and Mills explored the idea of constructing a Lagrangian density invariant under both local U(1) and local SU(2) transformations. Yang-Mills theories are based on the non-abelian SU(N) symmetry group and have many subtle and surprising properties. For instance, the two-component field,

$$\Psi = \begin{pmatrix} \psi_A \\ \psi_B \end{pmatrix}, \tag{1.11}$$

is said to have an internal degree of freedom, such that the two fields can transform into one another by an unitary SU(2) transformation. The elements of the SU(2) group can be written

$$U = e^{-i\alpha^k \tau^k}, \tag{1.12}$$

where $\alpha^k(x)$ are real parameters and τ^k are the three generators of the SU(2) group. From each τ^k there arises a gauge field, and in addition to the U(1) field, there is a total of four gauge fields.

In the 1960s Glashow, Salam and Weinberg attempted to couple the fermion fields with the new gauge boson fields W and Z. What resulted was a theory invariant under $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ transformations. This symmetry conserves weak *hypercharge* Y and the third component of weak *isospin* I_3. They are related by

$$Y = 2(Q - I_3), \tag{1.13}$$

where weak isospin is the internal degree of freedom shared by the members of an SU(2) multiplet. The members of a isospin multiplet carry the same weak hypercharge, but have different weak isospin. Only left-handed fermions participate in the weak interactions, thence the subscript L. For that reason, only left-handed fermions appear as doublets carrying internal isospin symmetry, while the right-handed leptons appear as singlets. Hence, charged right-handed[1] leptons only couple to the photon field. Linear combinations of the four gauge fields are interpreted as the real fields; the photon, the $\mathrm{W}^{\pm}$ and the Z fields. The non-abelian nature of the SU(2) group provides the gauge fields with the self-coupling ability.

However, electroweak theory is not a true unification of the electromagnetic and weak force as it has two coupling constants. They are related by the Weinberg angle, which must be determined experimentally.

[1] Right-handed neutrinos, as they have no charge, only interact through gravity. Even if they exist, they would hardly interact with matter at all.

Symmetry breaking

Initially, all fermions and gauge bosons were introduced as massless fields. However, most particles are observed to be massive, except from the photon and the gluon. As in the case for QED, introducing a mass term to the photon field destroys the gauge invariance, a problem shared by all fields in the SM. Yet another problem which arises in attempts to introduce mass terms, is non-renormalizable quantum theories yielding senseless infinite integrals.

A mass-generating mechanism retaining the gauge invariance of the Lagrangian density was, more or less independently, provided by Englert and Brout [4], Higgs [5][6] and Guralnik, Hagen and Kibble [7] in the 1960s. Renormalizability of massive Yang-Mills fields was later proven by Gerard 't Hooft [8]. If the lowest energy level in QFT, vacuum, is degenerate, an arbitrary state cannot be chosen as the ground state. Such a state will no longer share the symmetries of the Lagrangian, and *spontaneous symmetry breaking* has occurred. The staggering idea that the vacuum is non-unique, implies that some quantity is not invariant under symmetry transformations.

Consider a simple Lagrangian density respecting local U(1)×SU(2) symmetry,

$$\mathcal{L}_\Phi = (D_\mu \Phi)^\dagger D_\mu \Phi - V(\Phi^\dagger \Phi), \tag{1.14}$$

where Φ is a complex two-component field. Instead of defining the potential as $m^2 \Phi^\dagger \Phi$, such that $\Phi = 0$ corresponds to vacuum, an additional constant term is added,

$$V(\Phi^\dagger \Phi) = \frac{m^2}{2\phi_0^2} \left[\Phi^\dagger \Phi - \phi_0^2 \right]^2. \tag{1.15}$$

Thus the potential is shaped like a mexican hat with the minimum field energy located on a circle $|\Phi| = \phi_0$. In order to choose a suitable vacuum state as the ground state, one may utilize the three degrees of freedom of the SU(2) group, the real parameters $\alpha^k(x)$. Doing so, however, will spontaneously break the SU(2) symmetry. This freedom can be used to adopt a gauge where the ground state is given by

$$\Phi_{ground} = \begin{pmatrix} 0 \\ \phi_0 \end{pmatrix}, \tag{1.16}$$

while excited states are of the form

$$\Phi = \begin{pmatrix} 0 \\ \phi_0 + h(x)/\sqrt{2} \end{pmatrix}. \tag{1.17}$$

The *Higgs boson* is recognized as the quantum of the real scalar field $h(x)$. By interacting with the vacuum expectation value ϕ_0 of the Higgs field, the gauge bosons acquire their masses. Even though $SU(2)_L \times U(1)_Y$ symmetry has been broken, $U(1)_{EM}$ symmetry still remains, ensuring the photon a zero mass. It is the spontaneous symmetry breaking which is responsible for the electroweak force appearing as two separate forces; weak and electromagnetic.

Fermions acquire their masses through the *Yukawa* coupling, and the coupling constant is given by

$$\frac{m_f}{\sqrt{2}\phi_0}. \tag{1.18}$$

Hence, the Higgs boson couples more strongly to heavy fermions. The Yukawa couplings are however not given by theory, but must be determined by experiments.

1.1.2 Imperfections

Even though the SM has given a remarkably good description of nature in the energy range so far explored, it has its shortcomings. A so-called 'Theory of Everything' has to account for all physical phenomena, including constants, at all energy levels and times. Some of the flaws of the SM, motivating searches for new physics beyond the SM, are listed in the following.

- The SM, as it is based on QFT, is not compatible with general relativity. Hence, gravity is excluded as one of the fundamental forces. At energies well below the Planck scale $M_P \sim 10^{19}$ GeV the effect of gravity is negligible, but as the Planck scale is approached, a theory of quantum gravity is required.

- The SM contains 26 free parameters (25 if neutrinos correspond to Majorana fields[2]) which values are unrelated and must be experimentally determined. Three of them account for the coupling constants, two for the mass of the Higgs boson and the vacuum expectation value of the Higgs field, three for the charged lepton masses, six for the quark masses, four parameters for quark mixing, three for the neutrinos masses as well as four (or three) parameters for neutrino mixing.

- The SM doesn't provide an answer to; why there are three fermion generations, why an asymmetry between matter and antimatter in the universe is observed, why the universe is isotropic and homogeneous on a large scale, why quarks are confined in the hadrons, why charge is quantized and so on...

- Stars and galaxies move in ways that suggest there must be more matter present than what can be seen [9]. Based on observations, this *dark matter* accounts for 23% of the mass-energy density of the observable Universe, whereas ordinary matter only accounts for 4.6% [10]. The only SM candidate for dark matter is the neutrino, but due to their tiny mass, they move at speeds close to the speed of light and never had the chance to congregate in clumps. What is needed is rather a *cold dark matter* candidate, which is stable and heavy.

- Neither does the SM explain why there are such huge differences in the fundamental parameters such as masses and couplings, also known as the *hierarchy problem*. While the electroweak mass scale is $M_{ew} \sim 100$ GeV, the scale of gravity is many orders of magnitude larger; $M_P \sim 10^{19}$ GeV. The hierarchy problem is due to quantum loop corrections, which in their presence make the scales unstable. A solution might be provided by *Supersymmetry*[3] which relates the SM particles to superpartners of the opposite species (boson - fermion), canceling out the diverging terms in the Lagrangian. Another way out is the introduction of extra dimensions, possibly bringing the world of quantum gravity down to the TeV scale.

1.2 Theories of extra dimensions

The Planck scale has usually been considered as the fundamental scale from which the weak scale can be derived. However, recently the alternative viewpoint has gained interest. It might as well be the weak scale which is fundamental and from which the *four-dimensional* Planck scale

[2]Majorana fields are described by real spinors in contrast to the complex Dirac fields. Physically this means that a particle is its own anti-particle.

[3]Supersymmetric extensions to the SM also provide candidates for cold dark matter in the form of the Lightest Supersymmetric Particle (LSP). Depending on the model, the LSP could be the lightest neutralino, the gravitino or the lightest sneutrino.

is derived. In later times, several scenarios solving the hierarchy problem by introducing extra dimensions and confining the SM to a brane in higher-dimensional space have emerged. These render possible a fundamental Planck scale as low as the TeV scale. If this is the case, LHC might have the privilege to explore the phenomena of quantum gravity, including microscopic black holes (BHs).

In addition to the four visible space-time dimensions x^μ, $D-4$ extra dimensions with metric g_{mn} can be imagined. These are hidden by their size being smaller than the Planck length. They are denoted with coordinates y^m such that the D-dimensional metric can be written

$$ds^2 = dx^\mu dx_\mu + g_{mn}(y)dy^m dy^n. \tag{1.19}$$

This can be generalized, still respecting the 4D Poincaré invariance, to

$$ds^2 = e^{2A(y)}\eta_{\mu\nu}dx^\mu dx^\nu + g_{mn}(y)dy^m dy^n. \tag{1.20}$$

The SM confined to a brane at $y = y_0$ will then have an action of the form

$$S_{SM} = \int d^4x e^{4A(y_o)}\mathcal{L}\left(e^{2A(y_0)}\eta_{\mu\nu}, \psi_i, m_i\right), \tag{1.21}$$

which implies that four-dimensional physics scales vary depending on the location within the extra dimensions. The four- and D-dimensional Planck scales are related by

$$M^2_{P(4)} = M^{D-2}_{P(D)}V_w, \tag{1.22}$$

where V_w is the volume of the warped space. Thus the hierarchy problem may be eliminated by assuming the warped space to be large and flat, as in the model proposed by Arkani-Hamed, Dimopoulos and Dvali (ADD) [11]. This will be studied in more detail in Section 1.2.1. Another possibility, which is explored in the Randall-Sundrum (RS) model [12], is the existence of only one extra dimension and a high warp factor $e^{2A} \gg 1$.

In the simplest case, the extra dimension is compactified on a circle of radius R. Demanding any field propagating through the extra dimension to satisfy periodicity under the translation

$$y_i \to y_i + 2\pi R \quad , \quad i = 1, \ldots, n, \tag{1.23}$$

the wavenumber of the field is restricted to discrete values given by

$$k_n = \frac{n}{R}. \tag{1.24}$$

For an observer on the brane, the set of allowed momenta in the extra dimension appears as a tower of massive states of the propagating field,

$$M_n^2 = M_V^2 + \frac{n^2}{R^2}, \tag{1.25}$$

where M_V is the true mass of the field. This set of mass states is often referred to as the *Kaluza-Klein (KK) tower*. The spacing between the KK modes is specified by the compactification radius. For the ADD model, with R being in the range of ∼nm to ∼fm, this would imply a spacing of the order of less than an eV to a few MeV.

There exist several conventions for the fundamental Planck scale, which is related to the D-dimensional Newton's constant[4], and some of these are listed below.

[4]Newton's gravitational constant is measured to be 6.67428×10^{-11} $\mathrm{m^3kg^{-1}s^{-2}}$. In a D-dimensional Universe, the units become $\mathrm{m^Dkg^{-1}s^{-2}}$

- Convention used by Giddings and Thomas [13], useful in quoting experimental bounds:

$$M_P^{D-2} = \frac{(2\pi)^{D-4}}{4\pi G_D} \tag{1.26}$$

- Convention of Giudice, Rattazzi and Wells [14]:

$$M_D^{D-2} = \frac{(2\pi)^{D-4}}{8\pi G_D} \tag{1.27}$$

- Convention of Dimopoulos and Landsberg [15]:

$$M_{DL}^{D-2} = \frac{1}{G_D} \tag{1.28}$$

The physics does not depend on them, but they play an important role in the BH production threshold. In the following theory, BHs will be described in the context of the ADD model, and thereby the convention of Dimopoulos and Landsberg (Eq. 1.28) falls as the most natural choice.

1.2.1 The ADD model

In 1998 Nima Arkani-Hamed, Savas Dimopoulos and Gia Dvali proposed a framework for solving the hierarchy problem. The electroweak scale, which is experimentally verified, is taken as the only fundamental short distance scale in nature and at which the gravitational and gauge interactions become united. Due to the existence of $n \geq 3$ toroidal large[5] (compared to the electroweak scale) extra dimensions, gravity is observed to be exceedingly weak on distances $\gtrsim$10 nm. As the graviton[6] is expected to move freely in the *bulk* (D-dimensional) space, the strength of gravity leaks into the extra hidden dimensions and thus appears weak in the four-dimensional space-time.

Consider two point masses m_1 and m_2 separated by a distance $r \ll R$, where R is the radius of the n extra compact spatial dimensions. In a D-dimensional metric, the gravitational potential is dictated by Gauss' law in $D \equiv n + 4$ dimensions,

$$V(r \ll R) \sim \frac{m_1 m_2}{M_{DL}^{n+2}} \frac{1}{r^{n+1}}. \tag{1.29}$$

If the distance between the masses is greater than R, the gravitational flux lines can not propagate in the extra dimensions, and the well-known $1/r$ potential is obtained,

$$V(r \gg R) \sim \frac{m_1 m_2}{M_{DL}^{n+2} R^n} \frac{1}{r}. \tag{1.30}$$

The four-dimensional and fundamental Planck masses are thus related by

$$M_{P(4)}^2 = M_{P(4+n)}^{n+2} R^n, \tag{1.31}$$

where R^n is recognized as the warped space V_w from Eq. 1.22. This relationship clearly shows that the existence of large extra dimensions can push the fundamental Planck mass down to the

[5]Originally the article proposed two or more extra dimensions, but the case of $n = 2$ has in later times been experimentally excluded.

[6]Being a spin-2 particle, the graviton is not confined to a brane as particles with spin 0, 1 or 1/2 are by quantum field theory.

electroweak scale. What results is a new hierarchy among dimensions. However, this is a stable hierarchy which avoids fine-tuning problems[7].

If $M_P \sim 1$ TeV is assumed, it can be derived from

$$R \sim 10^{\frac{32}{n}-17}\text{cm} \tag{1.32}$$

that at least three extra dimensions are needed. The case of $n = 1$ implies $R \sim 10^{13}$ m, which would show up as deviations from Newtonian gravity over solar system distances, and thus it is empirically excluded. By requiring at least three extra dimensions, their radii decrease to less than $\sim$ 10 nm. At such distances gravity has not been probed yet. Since no deviations from Newton's inverse square law at larger distances have been observed, it follows that the SM particles can not freely propagate in the n extra dimensions. They are thought to be localized on a four-dimensional sub-manifold with thickness M_{ew}^{-1}.

However, in a sufficiently hard collision, particles with wavelengths smaller than R could be produced and escape into the extra dimensions, carrying away energy. This implies a sharp upper limit to the transverse momentum observed in four dimensions $p_T \leq E_{esc} \geq M_{ew}$. Even though energy can be lost into the extra dimensions, any gauge charge must be conserved in four-dimensional space where the gauge fields can propagate. As the graviton is free to propagate in the D-dimensional bulk, production of a graviton may also result in lost energy. Another interesting feature is the possibility of the escaped particles to orbit the extra dimensions, periodically returning to the four-dimensional world with frequencies of $R^{-1} \sim 10^8$ Hz for $n = 3$. At every return to the brane they may collide with other particles.

1.2.2 Constraints on M_P from other experiments

By looking for direct emission of KK gravitons, e.g. in the $q\bar{q} \rightarrow g/\gamma + G_{KK}$ process which results in a monojet or monophoton + missing transverse energy, constraints on the fundamental Planck scale have been set. The case of $n = 2$ has recently been excluded by Cavendish-type experiments [16], but for a higher number of dimensions the current constraints from LEP [17] and Tevatron [18][19] on M_P are close to 1 TeV, see Table 1.3.

Experiment and channel	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$
LEP combined	1.20	0.94	0.77	0.66	-	-
CDF monophotons, 2.0 fb^{-1}	1.00	0.97	0.93	0.90	-	-
DØ monophotons, 2.7 fb^{-1}	0.90	0.87	0.85	0.83	0.83	0.80
CDF monojets, 1.1 fb^{-1}	1.08	0.98	0.91	0.88	-	-
CDF combined	1.15	1.04	0.98	0.94	-	-

Table 1.3: Most recent 95% CL lower limits on the fundamental Planck scale M_D (TeV).

1.3 Higher-dimensional black holes at the TeV scale

BH formation is a classic non-perturbative process not appearing in any order of perturbation theory. In other words, there are no small numbers or coupling constants, and the cross section

[7]When parameters of a model need highly precise adjustment to agree with observations, it is referred to as a fine-tuning problem.

grows with the center-of-mass energy $\sqrt{s}$ at a rate determined by the number of extra dimensions and their geometry,

$$\sigma \sim s^{1/(n+1)}. \tag{1.33}$$

If TeV scale gravity is reality, hadron colliders such as the Large Hadron Collider (see Section 2.1) will become BH factories with a BH production rate of up to ~ 1 Hz. Other hard perturbative processes will be highly suppressed for $\sqrt{s} \gg M_P$. This can be understood by the event horizon forming before the incident particles come in causal contact, such that any hard processes are cloaked behind the horizon. In other words, particles produced in the collisions can not escape the scattering center.

BHs are expected to decay thermally by Hawking evaporation, giving rise to spectacular spherically distributed events with high multiplicity. Triggering is easy, and no special purpose trigger is needed since the BH events will contain at least one prompt lepton or photon with energy > 100 GeV and numerous energetic jets[8]. These features provide a clean signature with low background. BHs would in fact represent a significant background in many new physics searches if TeV scale gravity is realized.

The study of higher-dimensional BHs and the correlations between its mass and temperature can test the validity of Hawking's evaporation law, determine the number of large extra dimensions and the fundamental Planck scale. When quantum gravity is probed, new symmetries which guarantee proton longevity and approximate conservation of lepton and flavour numbers may be revealed.

The material presented in the following section is drawn from the article on BHs at the LHC by S. Dimopoulos and G. Landsberg [15] and a more thorough review by S. Giddings and S. Thomas [13]. Nevertheless, there are numerous texts available on the topic TeV-scale BHs, and the interested reader is referred to the reviews [20][21][22][23], just to list a few.

1.3.1 Thermal black holes

BHs far exceeding the fundamental Planck mass are well-understood entities in the treatment of general relativity. When their masses approach the Planck mass, however, their properties become complex and model dependent, and a full theory of quantum gravity is required. In order to describe tiny BHs, it is common to ignore this obstacle and estimate their properties by semi-classical[9] arguments. These are strictly speaking only valid for $M_{BH} \gg M_P$. A stringent criterion for whether a semi-classical description can be trusted is given by the number of degrees of freedom. This is roughly equivalent to the entropy of the BH, and small statistical fluctuations thus require

$$\sqrt{S_{BH}} = \sqrt{\frac{kA}{4\ell_P^2}} \gg 1. \tag{1.34}$$

Here, k is Boltzmann's constant, A is the area of the event horizon, and ℓ_P is the Planck length. In the case of $D = 10$ and $M_P = 1$ TeV, BHs with masses $5M_P$ and $10M_P$ would give $\sqrt{S_{BH}} \simeq 5$ and $\sqrt{S_{BH}} \simeq 8$, respectively. Clearly, a semi-classical treatment is not valid for BHs with masses beneath $10M_P$, but in lack of a better theory, this is the approximation used. However, the approach relies on some properties of the BH to be fullfilled.

Firstly, the Schwarzschild radius of the BH is required to be small compared to the radii of the extra dimensions and the scale on which the warp factor is employed. Periodic boundary

[8] This applies to *thermal* BHs, which will be discussed in Section 1.3.1

[9] Semi-classical gravity treats SM fields as quantum fields and the gravitational field as a classical field. This is an approximate theory of quantum gravity.

conditions can then be neglected and space-time treated as spherically symmetric. The metric of a Schwarzschild BH is given by

$$ds^2 = -\left[1 - \left(\frac{R_S}{r}\right)^{n+1}\right] dt^2 + \frac{1}{1 - \left(\frac{R_S}{r}\right)^{n+1}} dr^2 + d\Omega_{3+n}^2, \tag{1.35}$$

with $d\Omega_{3+n}$ being the surface element of a $(3+n)$-dimensional sphere and R_S the Schwarzschild radius of the BH.

Secondly, the BH must have a mass higher than the tension of the three-brane on which the SM-particles live. If so, gravitational effects can be neglected, and the only effect of the brane field is to bind the BH to the brane. This is called the 'probe brane approximation'. Other fields from string theory are assumed not to play an important role.

Production

The Hoop Conjecture proposed by Kip Thorne in 1972 [24] states that black holes only form when an object is compacted into a region whose circumference in all directions is less than $2\pi R_S$, where R_S is the Schwarzschild radius of the object's mass. From this the cross section for BH production in parton-parton scattering is expected to be of the form

$$\sigma_{ij\to BH} \sim F(s)\pi R_S^2(\sqrt{s}), \tag{1.36}$$

where $F(s)$ is a dimensionless form factor accounting for the fact that not all of the available energy is captured behind the event horizon[10].

The effect of gauge charges is assumed to be negligible, but spin has to be taken into consideration. Neutral spinning BHs are described by the higher-dimensional *Kerr* solutions, which are discussed by Myers and Perry in [27]. The Schwarzschild radius for a Kerr BH in the ADD framework is given by

$$R_S = \frac{1}{\sqrt{\pi} M_{DL}} \left[\frac{M_{BH} 8\Gamma\left(\frac{n+3}{2}\right)}{M_{DL}(n+2)}\right]^{1/(n+1)}. \tag{1.37}$$

Inserting this into Eq. 1.36 yields

$$\sigma_{ij\to BH}(s) = F(s)\frac{1}{M_{DL}^2} \left[\frac{M_{BH} 8\Gamma\left(\frac{n+3}{2}\right)}{M_{DL}(n+2)}\right]^{2/(n+1)}, \tag{1.38}$$

which is valid in the semi-classical limit $\sqrt{s} \gg M_P$. This indicates that the cross section increases with the mass of the black hole ($\sim \sqrt{s}$) at a rate decreasing with the number of extra dimensions. The decreasing rate is a result of the gravitational field spreading out in the space of extra dimensions. Hence, a larger number of dimensions allows for more energy to escape into the bulk. It is clear that the discovery potential is highly dependent on the Planck scale being low.

Having the Hoop Conjecture in mind, a naive estimate of the production threshold can be calculated. The Compton wavelength in the rest frame of the two colliding partons with energy $\sim E/2$ must lie within the Schwarzschild radius of a BH with energy $E \sim M_{BH}$. In natural units the Compton wavelength is simply given by the inverse of the energy

$$\lambda = \frac{1}{E}. \tag{1.39}$$

[10]Estimates of the form factor based on Schwarzschild-Tangherlini BHs, neglecting spin and charge effects, are given in [25]. In all higher-dimensional cases the form factor is $\gtrsim 1$ and increases with D. Moreover, further improvements of estimates [26] even raise the cross section by a significant amount in the range 40%-70%.

Therefore, $R_S > 2/E$ must be required, giving a lower bound on the mass of the BH. Figure 1.1 shows $x_{min} = M_{BH}^{min}/M_P$ as a function of n calculated by using the semi-classical formula Eq. 1.37. Again, it must be emphasized that this is not valid in the Planck-regime $M_{BH} \sim M_P$, but at least it gives an indication of where the actual production threshold might be. With the M_D Planck scale convention (Eq. 1.27), Meade and Randall found $x_{min} > 0.44$ for $n = 6$ [28].

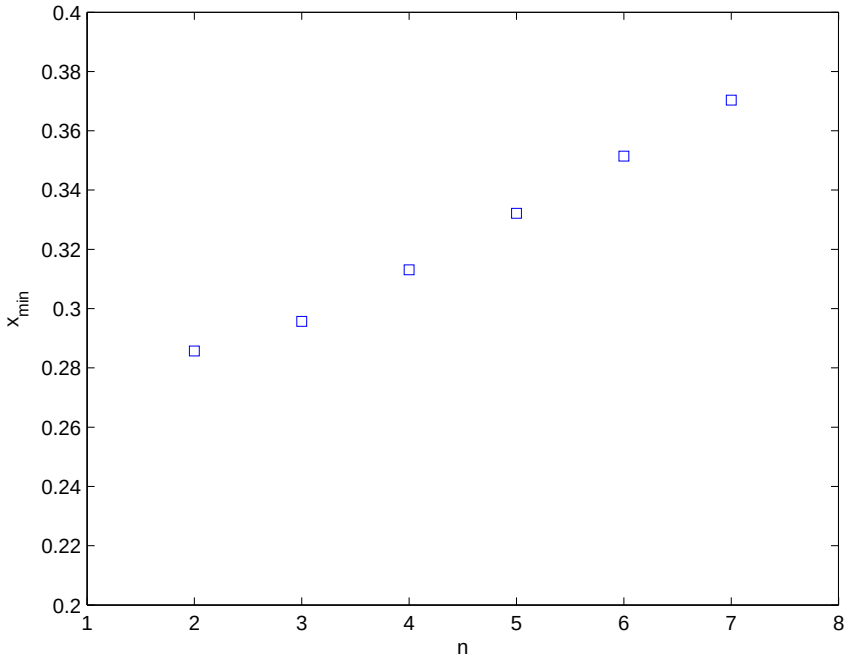


Figure 1.1: Naive estimate of $x_{min} = M_{BH}^{min}/M_P$ as a function of the number of extra dimensions n for semi-classical BHs.

To obtain the total $pp \to BH$ cross section, Eq. 1.38 must be convoluted with the parton distribution functions, and a sum over all possible parton pairings is needed. Hence,

$$\sigma_{pp\to BH}(\tau_m, s) = \sum_{ij} \int_{\tau_m}^{1} d\tau \int_{\tau}^{1} \frac{dx}{x} f_i(x) f_j(\tau/x) \sigma_{ij\to BH}(\tau s), \tag{1.40}$$

where x is the parton momentum fraction, $\tau = x_i x_j$ is the parton-parton center-of-mass energy squared fraction, $\sqrt{\tau_m s}$ is the minimum center-of-mass energy for which Eq. 1.38 is applicable, and the BH mass is assumed to be $\simeq \sqrt{\tau s}$.

Decay

The decay of a semi-classical BH is thermal and governed by its Hawking temperature,

$$T_H = \frac{n+1}{4\pi R_S}, \tag{1.41}$$

from which it can be extracted that the BH gets hotter as its mass decreases. The change in entropy w.r.t. the change in mass is given by

$$\frac{\partial S_{BH}}{\partial M_{BH}} = \frac{1}{T_H}, \tag{1.42}$$

and solving for the entropy gives

$$S_{BH} = \frac{n+1}{n+2}\frac{M_{BH}}{T_H}. \tag{1.43}$$

A small BH thus has less entropy than a larger one. The average multiplicity is given by

$$\langle N \rangle \approx \frac{M_{BH}}{2T_H}, \tag{1.44}$$

which implies that small, and thus very hot, BHs emit a small number of highly energetic particles. All conservation laws are obeyed in thermal decays, and they don't discriminate between particle species.

The decay proceeds through several stages, each with its characteristic energy spectrum, multiplicity and angular distribution.

- The first phase is called the *balding phase*. The initial configuration of the BH has a highly asymmetric horizon and a possibly non-trivial topology. By emitting SM particles and gravitons it sheds its "hair", causing the horizon to grow and become symmetric. The BH is now a spinning Kerr solution and it is estimated by D'Eath [29] to have lost $\sim 16\%$ of its energy. Any gauge charges present have been carried away by the radiation.

- Second is the brief *spin-down phase*, where semi-classical Hawking radiation strips the BH of its angular momenta by emitting quanta predominantly in the $l = m \sim 1$ modes. Such particles are somewhat more likely to be emitted in a direction perpendicular to the spin axis of the BH, but still fairly spherical. The particles possess characteristic energies determined by the Hawking temperature T_H. Roughly 25% of a four-dimensional BH's energy will be radiated in this phase, and a similar fraction of energy-loss is expected in the higher-dimensional case.

- Next is the longer *Schwarzschild phase*, in which the black hole continues to emit Hawking radiation with a thermal spectrum determined by T_H. The spin-down and Schwarzschild phase are also collectively called the Hawking evaporation phase. After spinning down, the BH acts as a point source which radiates s-wave particles uniformly distributed on a sphere. Hence the angular distribution of the particles emitted during the evaporation phase is given by

  $$\frac{dN}{d\phi} \sim N_0 + 2N_1 sin^2\phi, \tag{1.45}$$

  where N_0 and N_1 are the number of particles emitted in the Schwarzschild and spin-down phases, respectively, and ϕ is the angle with respect to the spin axis in the rest frame. The temperature of the BH rises until it reaches a mass $\sim M_P$ where Hawking's equations are no longer valid.

- The last phase is the *Planck phase*, which in the absence of a theory of quantum gravity, is poorly understood. It is however expected that the BH either completely decays in emission of a few highly energetic quanta or leaves a stable remnant.

In [30] it is shown that the rate at which a (n+4)-dimensional field loses energy is only dependent on a scale. Hence, the BH should decay with equal probability to any degree of freedom on either the brane or in the bulk. Since there are many more particle species living on the brane than in the bulk, the BH would mainly decay to visible SM particles.

Accounting for decay of top quarks, massive gauge bosons and the Higgs boson, a ratio of roughly 5:1 of hadronic to leptonic activity is expected in the evaporation phases. Being dependent on the intrinsic spin of the particle and ignoring particle masses, the fractions of

quarks and gluons, charged leptons, massive gauge bosons, invisible (neutrinos and gravitons), Higgs bosons and photons emitted from a non-rotating four-dimensional black hole are estimated to be 72%, 11%, 8%, 6%, 2% and 1%, respectively. These numbers are expected to be indicative for higher-dimensional BHs, and accounting for e.g. supersymmetric particles would of course modify the ratios. For the higher-dimensional BH, it is expected that 85-90% of the energy is visible, making measurements of its mass quite precise.

Moreover, due to the rapidly falling structure functions of the partons contained in the proton, BHs are typically produced with a moderate boost $\langle\gamma\beta\rangle \lesssim 1$ in the laboratory frame. Thus the decay products are not highly boosted, contributing to a high degree of sphericity. A completely spherical event would correspond to the decay of a non-spinning BH with zero boost, which would result in the transverse energy being 50% of the total energy. When accounting for spin and boosting, the transverse energy will tend to account for between $\frac{1}{3}$ and $\frac{1}{2}$ of the total energy.

1.3.2 Searching for black holes with ATLAS during the low luminosity phase

Unfortunately, multi-particle final states are suppressed since only BHs produced well above threshold have sufficient entropy to decay thermally, and this is unlikely to happen at the LHC [28], at least in the low luminosity phase. However, there is no need to despair, as truly thermal BHs wouldn't provide any insight to quantum gravity, except from verifying already existing theories. Due to the rapidly falling parton density functions (PDFs) and inelasticity[11], most BHs would rather be produced at or just above threshold. It is however expected that even the low multiplicity final states will be measured as much more transverse than the background, and BHs produced just above threshold will produce the hardest radiation due to their high temperature. Since these BHs have masses of the order of the Planck mass, they are hereafter referred to as Planckian BHs (PBHs)[12]. It is believed that PBHs, being true or virtual BHs or simply strong gravity effects, will lead to an increase in the $2 \to 2$ production cross section when the center-of-mass energy approaches the Planck scale.

It is assumed that PBHs are defined by three quantities; mass, spin and gauge charge. Local gauge symmetries are presumably not violated by gravity, such that the charges of the SM are conserved. However, as the world of quantum gravity is unfolding, its effects will play an important role, and a most dramatic possible signature is final states violating global quantum numbers such as e.g. baryon and lepton number. Thus PBHs could decay to final states with small or no SM background, such as lepton + jet.

Another possibility for PBH decay is the formation of a stable remnant [31][32][33], an idea motivated by several considerations, e.g. the information loss puzzle. Some of the associated signatures are delayed flashes of hard photons, leptons or mono-jets and the complete disappearance of high p_T back-to-back di-jets above the BH production threshold $\sim M_P$.

Although the scenario of BH remnants is highly interesting, it relies on exotic assumptions. Since PBHs are in most scenarios expected to decay to low-multiplicity states, preferably two particles, a plausible search strategy is to look for signatures in the di-lepton channel.

Di-lepton final states

Pure lepton final states are known for being the most robust channels for analysis due to their clean signatures, easy triggering and simple reconstruction. Since the LHC collides protons, the two colliding partons are most likely to be quark and gluon or two quarks, i.e. produced PBHs will most likely carry colour charge. Assuming that local gauge symmetries are conserved by

[11] The colliding partons lose energy before being fully trapped inside the event horizon.

[12] The term PBH in this thesis always refers to a two-particle final state BH.

quantum gravity, a di-lepton final state can only result from a quark-antiquark or gluon-gluon collision. However, with the lepton background being smaller than the jet background, the di-lepton signal should still dominate over the background. Due to the clean signatures of leptons, the di-lepton channel will, if a signal is detected, be important in distinguishing among quantum gravity models, a task which is beyond the scope of this thesis.

As mentioned earlier, it is reasonable to expect the $2 \rightarrow 2$ cross section to increase when PBH production comes into play. The very hot PBHs do not go through the process of balding, spinning down and emitting Hawking radiation before the final explosion (as described in Section 1.3.1), but decay instantaneously. The average number of decay products $\langle N \rangle$ can thus be approximated by a Poisson distribution, such that the PBH $\rightarrow 2$ cross section is given by

$$\sigma_{PBH \rightarrow 2} = \pi r_S^2 e^{-\langle N \rangle} \sum_{i=0}^{2} \frac{\langle N \rangle^i}{i!}. \tag{1.46}$$

In the non-perturbative regime, it is not known how the PBH cross section will interfere with the SM cross section. Due to this uncertainty, the analysis in this thesis will be built on the assumption that the cross sections can simply be added.

It has been claimed that even though BHs preserve information, any global quantum numbers following from an effective or anomalous symmetry are violated [34]. The SM Lagrangian possesses several global U(1) symmetries, such as baryon number B and lepton number L. At low energies, these symmetries play an important role, but it is widely believed that they are broken at higher energies. Extending to SU(2)×U(1), both B and L are anomalous symmetries conserved by perturbative processes, but violated in non-perturbative ones. However, B - L remains conserved within the SM.

At the low energies probed so far, the proton has not been observed to decay. The proton consists of two up quarks, the lightest quark, and one down quark. Due to energy conservation, the down quark is prohibited to decay to an up quark. This might sound strange, but the lightest composite particle consisting of three up quarks, also known as $\Delta^{++}(1232)$, has a larger mass than the proton. This is due to the Pauli exclusion principle, which states that the wave-function for two identical fermions is anti-symmetric. Since the colour part of the wave function for the $\Delta^{++}(1232)$ is necessarily anti-symmetric, the spin part must be symmetric. All quark spins are thus aligned in the same direction, resulting in an exited spin state $S = 3/2$. Violating both baryon and lepton number, but conserving B - L, a possible proton decay channel is

$$p^+ \rightarrow e^+ + \pi^0. \tag{1.47}$$

In the context of any grand unified theory (GUT), quarks and leptons of a given family would be members of the same representation of the GUT gauge group. Hence, B and L violating processes are expected to be generic. The captured matter is compressed inside a BH, and when reaching the GUT-scale densities, the system becomes baryon charge neutral. This mechanism should be effective for long-lived macroscopic BHs, but there is no definite solution for virtual BHs without fully understanding quantum gravity. There are, however, strong indications that baryon charge is not conserved within virtual BHs as well.

In principle, all charge conservation laws stemming from effective or low-energy global symmetries can be violated by information-preserving (unitary) BHs. This applies to, in addition to baryon number B, all fermion flavours, lepton number L, charge conjugation C, parity P and the combined symmetry CP. Many GUTs conserve B - L, so this quantity might well be conserved within BHs. However, CPT symmetry is required by any self-consistent local field theory and is expected to be conserved.

Opposite-sign same-flavour (OSSF) di-leptons are the only possible di-lepton final state in the case where all lepton flavours are conserved. If not, opposite-sign opposite-flavour (OSOF)

(e.g. $e\mu$ and $\mu\tau$) and OSSF di-leptons final states should occur at the same rate. Being restricted to two final state particles, turning off B and L conservation will not increase the direct di-lepton decay rates. This is due to colour and electrical charge conservation, such that di-lepton final states can still only result from neutral gluon-gluon or quark-antiquark collisions.

Chapter 2

Apparatus

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the world's largest particle accelerator and is designed to collide both protons and heavy ions. It is situated 50 to 175 meters below both French and Swiss ground in the Geneva area. Being 27 kilometers in circumference, 1232 superconducting dipole magnets operating at 8.3 Tesla are needed to bend the 7 TeV particle beams. At full intensity, each proton beam will consist of 2808 bunches, each containing 1.15×10^{11} protons, giving 600[1] million inelastic events per second at a nominal luminosity 10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. When fully operating, the LHC will accelerate protons to an energy of 7 TeV giving a center of mass energy of 14 TeV.

The protons are obtained by stripping hydrogen atoms of electrons and are prepared by a series of systems aiming to increase their energy. They are first injected from the linear accelerator LINAC2 at 50 MeV into the Proton Synchrotron Booster (PSB), where the photons are accelerated to 1.4 GeV. Then they are sent to the Proton Synchrotron (PS), which gives the protons an energy of 25 GeV before they arrive to the Super Proton Synchrotron (SPS). Here their energy increases to 450 GeV before the journey ends in the LHC, in which they will circulate for 20 minutes until reaching maximum energy. The accelerator complex is illustrated in Figure 2.1.

On the 10th of September 2008, the first beam was circulated through the LHC. Unfortunately, 9 days later, a quench occurred in some of the bending magnets causing a massive leak of liquid helium into the tunnel. This led to a total of 53 damaged magnets, and not until November 2009 was the LHC again operational with collisions on the 23rd. Achieving 1.18 TeV per beam on the 30th of November, the LHC became the highest-energy particle accelerator in the world. After a technical stop, the LHC was up and running at 3.5 TeV per beam in March 2010. At the end of the proton running in 2010, the major target of a collision rate of 10^{32} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ was achieved. The LHC will continue the proton running throughout 2011, with 3.5 TeV per beam and an expected peak luminosity of 1.3 - 1.8 $\times 10^{33}$ $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The integrated luminosity of the 2011 running is expected to be 2.7 - 3.7 fb^{-1}, and as of June 15, 2011 ATLAS had collected 1 fb^{-1} of data.

In total there are six experiments at the LHC. ATLAS and CMS are general-purpose detectors for analysing the whole spectra of particles produced in the collisions. ALICE and LHCb are specialized for analyzing specific phenomena, e.g. quark-gluon plasma and studies of the beauty quark. Finally there are TOTEM and LHCf. TOTEM is designed to focus on elastic collisions in order to measure the poorly understood internal structure of the proton. It also

[1]The design bunch crossing rate is indeed 40 millions per second, but one expects several particle collisions per crossing.

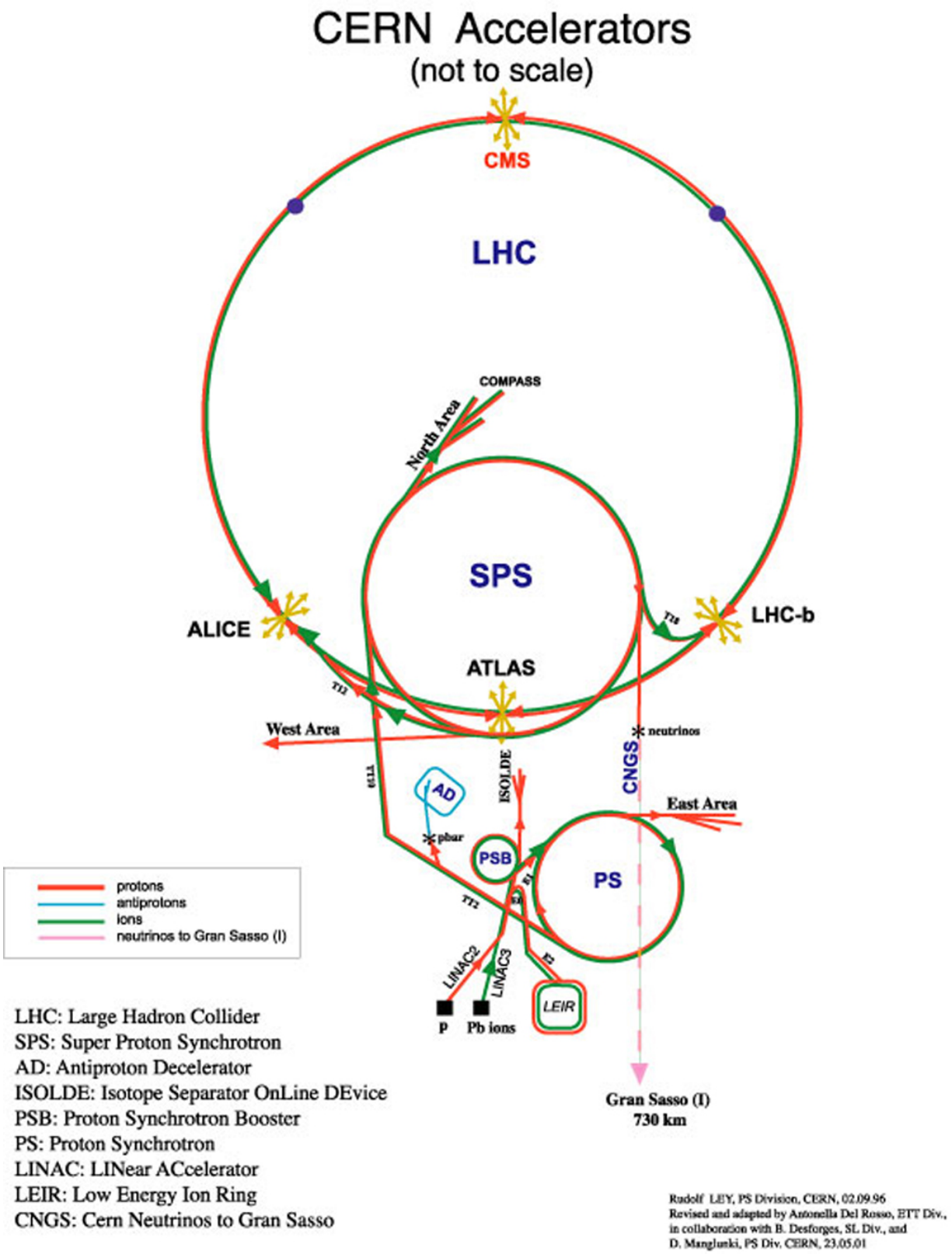


Figure 2.1: The accelerator complex at CERN. Illustration from CERN Document Server.

gives accurate measurements of the luminosity at LHC. LHCf is used to understand forward

particle production in very small angle pp-collisions for calibration of large-scale cosmic-ray experiments. This thesis will analyze data from the ATLAS detector, of which a brief discussion of its properties will follow.

2.2 The ATLAS detector

The ATLAS (**A** **T**oroidal **L**HC **A**pparatu**S**) detector [35] is designed to detect the products of a proton-proton collision. As a general-purpose detector, it will investigate a wide range of physics topics and might even unveil something new and unexpected. Some of the main goals for the ATLAS detector are the discovery of the Higgs boson and supersymmetric particles, including a possible candidate for the puzzling dark matter. The high center of mass energy and luminosity provided by LHC will produce a great number of heavy top quarks, allowing ATLAS to make more precise measurements of its mass and interactions. One of the most exciting possible discoveries, however, is microscopic black holes, which would imply extra dimensions.

With its 7000 tonnes, 44 meters in length and 25 meters in diameter, ATLAS is the largest-volume particle detector ever built. It contains six detecting subsystems for identifying particles and measuring their energy and momentum and a huge superconducting magnet system for bending charged particles, see Figure 2.2. With up to 600 million events per second, the ATLAS detector will generate an impressive amount of data. A highly advanced trigger system is needed to identify the most interesting events. What follow is a quick overview of the main components of ATLAS.

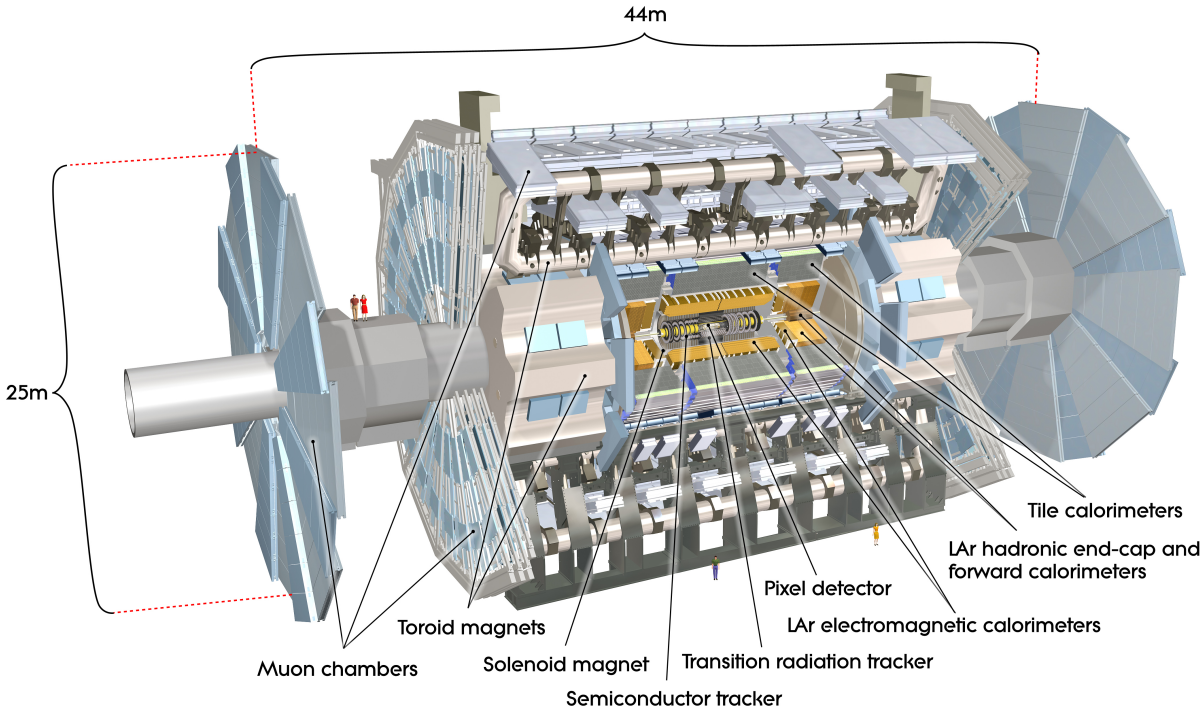


Figure 2.2: An overview of the ATLAS detector. Illustration from CERN Document Server.

2.2.1 Coordinate system

The nominal interaction point is defined as the origin of the coordinate system with the beam moving along the z-direction, transverse to the $x - y$ plane. The positive x-axis points from the interaction point to the center of the LHC ring, while the positive y-axis points upward. The azimuthal angle ϕ is measured around the beam axis, while the polar angle θ is measured as the angle from the beam axis. A widely used quantity is the pseudorapidity, defined as $\eta = -\ln \tan(\theta/2)$. The transverse momenta p_T and energy E_T are defined in the x-y plane, transverse to the beam axis. Finally, the distance ΔR in the $\eta - \phi$ space is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

2.2.2 Magnet system

The magnet system consists of one solenoid surrounding the inner detector (ID) and three air-core toroids, one barrel and two end-caps. Forming the basis of the muon system, the toroids are arranged with an eight-fold azimuthal coil symmetry placed around the calorimeters.

As the solenoid is situated in front of the electromagnetic calorimeter, its layout is carefully optimized to keep the thickness as low as possible, about ∼0.66 radiation lengths. As a result of a compromise, the solenoid is shorter than the ID cavity. A shorter coil reduces the material in front of the calorimeter, while a longer coil gives a more uniform magnetic field in the ID. To improve the performance of the ECAL even further, the central solenoid and the EM barrel calorimeter share the same vacuum vessel, eliminating two vacuum walls. Accuracy of the momentum scale is essential in the ID cavity, and thus the field therein has been mapped for different solenoid currents at 20,000 space points. The solenoid provides 2 T at the interaction point, but this drops to about 0.9 T at the end of the ID cavity.

Each toroid consists of eight coils symmetrically assembled around the beam axis. In order to provide radial overlap, the end-cap coils are rotated 22.5° with respect to the barrel coils. The barrel toroid delivers a field varying from 0.15 T to 2.5 T, depending on R and azimuthal angle ϕ, with an average of 0.5 T. Similarly for the end-caps, they provide a field varying from 0.2 T to 3.5 T with an average of 1 T. In contrast to the solenoid field, the toroid field is highly non-uniform. It is continuously monitored by about 1800 Hall sensors in all directions.

2.2.3 Inner detector

The ID measures the trajectory of charged particles through their interaction with the material at discrete points. From inside out the ID consists of three detector types, the silicon Pixel detector, the semiConductor Tracker (SCT) and the Transition Radiation Tracker (TRT), see Figure 2.3. Due to the very high density of tracks, it is essential that the detectors have fine granularity and are able to do high-precision measurements. The ID system is designed to provide excellent momentum resolution for charged particle tracks with p_T as low as 0.5 GeV, measurements of both primary and secondary vertices, as well as electron identification for $|\eta| < 2.0$. In Table 2.1 the intrinsic accuracies of the ID sub-systems are displayed.

	Pixel			SCT		TRT
	Layer-0	Layer-1 and -2	Disks	Barrel	Disks	
Intrinsic accuracy (μm)	10 (R-ϕ) 115 (z)	10 (R-ϕ) 115 (z)	10 (R-ϕ) 115 (R)	17 (R-ϕ) 580 (z)	17 (R-ϕ) 580 (R)	130

Table 2.1: Intrinsic measurement accuracies for the ID sub-systems. The numbers correspond to the single-module accuracy for the pixels, the effective single-module accuracy for the SCT and the drift-time accuracy of a single straw for the TRT.

The Pixel detector and SemiConductor Tracker

The precision tracking detectors, the Pixels and the SemiConductor Tracker (SCT), cover the region $|\eta| < 2.5$. They are arranged in concentric cylinders around the beam axis together with disks located at each end-cap region. Both are based on the semiconducting crystalline material silicon. When charged particles pass through the material, electron-hole pairs are created. These are collected by an electric field giving a small current. As silicon detectors only require a few eV to create a pair, the energy resolution provided is unmatched by any other detector type. The silicon detectors allow for measurements of impact parameter and secondary vertices for heavy-quark and τ-lepton tagging.

Using pixel detectors closest to the vertex, the highest granularity is achieved. The pixel detectors contribute mainly to the accurate measurements of vertices and typically provide three space-point measurements. They consist of a barrel part with three cylindrical layers and three disks on each side, perpendicular to the beam axis. The pixel sensors are 250 μm thick oxygenated n-type silicon wafers with 46,080 pixels, each sized 50×400 microns in $R - \phi \times z$. In total there are 1744 pixel sensors, giving a staggering number of 80.4 million pixels. The readout is on the n^+-implanted side of the detector, allowing for good charge-collection efficiency after type inversion [2] and lower depletion voltage. Unfortunately, the ID is exposed to high radiation causing severe damage. Therefore all components are selected for their radiation hardness. However, it is planned for that the innermost pixel layer must be replaced roughly every third year. The overall radial extension of the pixel detector is 242 mm.

The semiConductor Tracker (SCT) uses classic single-sided p-in-n microstrips for a more practical and economical coverage of a larger area. There are 15,912 sensors with a length of 12 cm and a thickness of 285$\pm$15 μm. Each sensor consists of 768 microstrips with a pitch of 80 μm. By arranging the strips in double layers, one along the beam direction (in the barrel) or radial (in the end-caps) and one with a stereo-angle of 40 mrad, the SCT is capable of measuring both $R - \phi$ and z or R coordinates. In total, the four barrel double-layers and 18 end-cap disks sum to 6.3 million readout channels. Typically, the SCT modules provide four precision space-point measurements for each track. Including the pixel detector, the precision tracking detectors have a radial extension of 549 mm for the barrel and 610 mm for the end-cap.

The Transition Radiation Tracker

The TRT is the outermost detector of the ID system. It has a larger radial coverage, resulting in a large number of measurements and longer measured track length which compensates for the lower precision. It consists of numerous straw tubes with a diameter of 4 mm filled with a gas mixture of $Xe/CO_2/O_2$. When a charged particle passes through the gas, it will create ion-electron pairs (ionization) giving rise to a current when voltage is applied.

Interleaved between the straws are polypropylene-foils serving as the transition radiation material. Transition radiation, photons emitted by charged particles crossing the interface of two media with different dielectric constants, causes more ionization of the gas. Since the amount of transition radiation is greater the higher the Lorentz factor of the particle, particle tracks with many large signals can be identified as the lightest charged particle, namely the electron.

The barrel part is divided into three rings of 32 modules each, where each module consists of 52,544 straws, while the end-caps have 20 modules each with 122,880 straws. In total, the TRT has approximately 351,000 readout channels, measuring drift time and giving a spatial resolution of 130 μm. On average the TRT provides 36 tracking points, which enables track-following up to $|\eta| = 2.0$. The TRT also has two independent thresholds, allowing it to discriminate between tracking hits and transition radiation hits.

[2] Due to radiation damage, the detector will over time change to a p-type semiconductor.

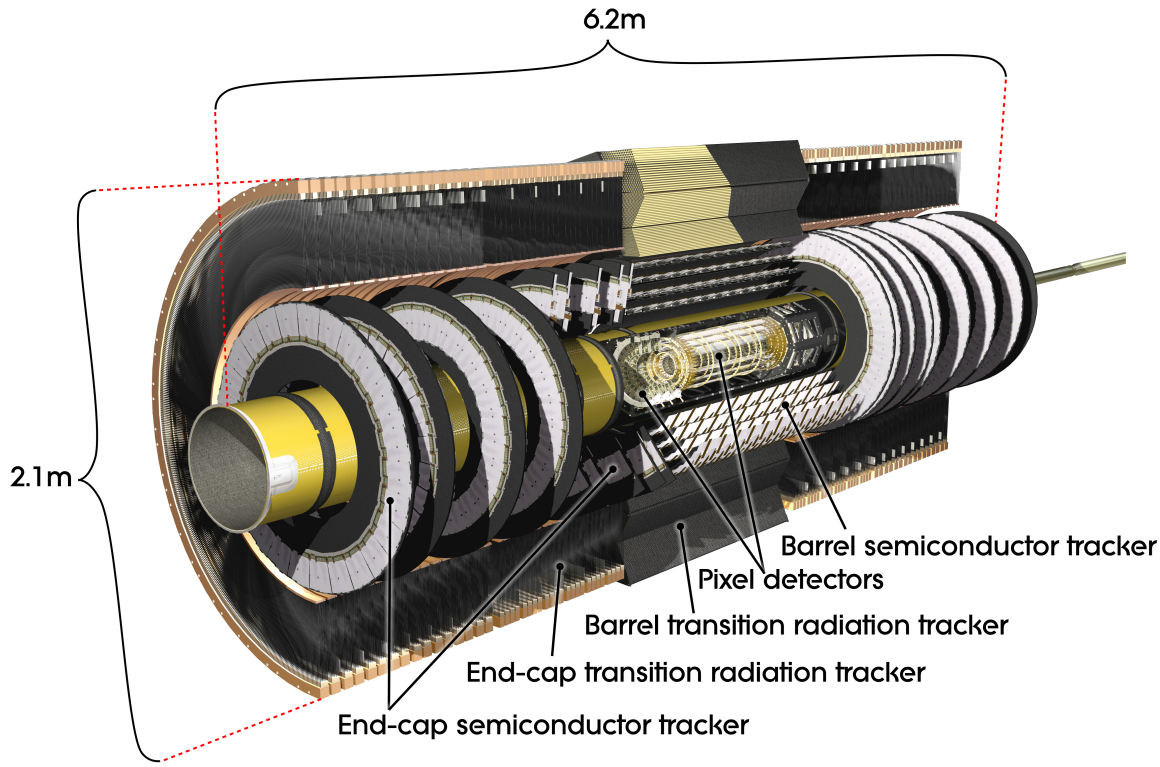


Figure 2.3: The inner detector layout. Illustration from CERN Document Server.

2.2.4 Calorimetry

Situated outside the inner detector and the solenoidal magnet, the calorimeters are designed to measure the energy of particles. The system is split in two; an inner electromagnetic calorimeter (ECAL) and an outer hadronic (HCAL) calorimeter. As the names suggest, the ECAL absorbs energy from particles which create interact electromagnetic showers, i.e. electrons and photons, while the hadronic calorimeter absorbs energy of strongly interacting particles such as pions and nucleons. Both are based on sampling techniques.

The innermost calorimeters are placed in three cryostats, one barrel and two end-caps. Whereas the barrel contains the ECAL barrel and the solenoid magnet, the end-caps house both the EM end-cap calorimeter (EMEC), the hadronic end-cap calorimeter (HEC) and a forward calorimeter (FCal), all of which use liquid argon as the ionization medium. Argon has been chosen due to its linear behaviour, stability of response over time and hardness against radiation.

The calorimetry of ATLAS, see Figure 2.4, covers the range $|\eta| < 4.9$ and has a thickness sufficient to stop most particles, apart from muons and neutrinos, from reaching the muon system. These properties ensures good jet reconstruction and measurements of missing transverse energy[3] E_T^{miss} important for a variety of physics signatures.

Electromagnetic calorimetry

The ECAL provides fine granularity for precise measurements of electrons and photons. This is important for detecting Higgs bosons decaying into electrons via W or Z-pairs or photons

[3]In practice all neutrinos escape ATLAS and carry away energy and momentum.

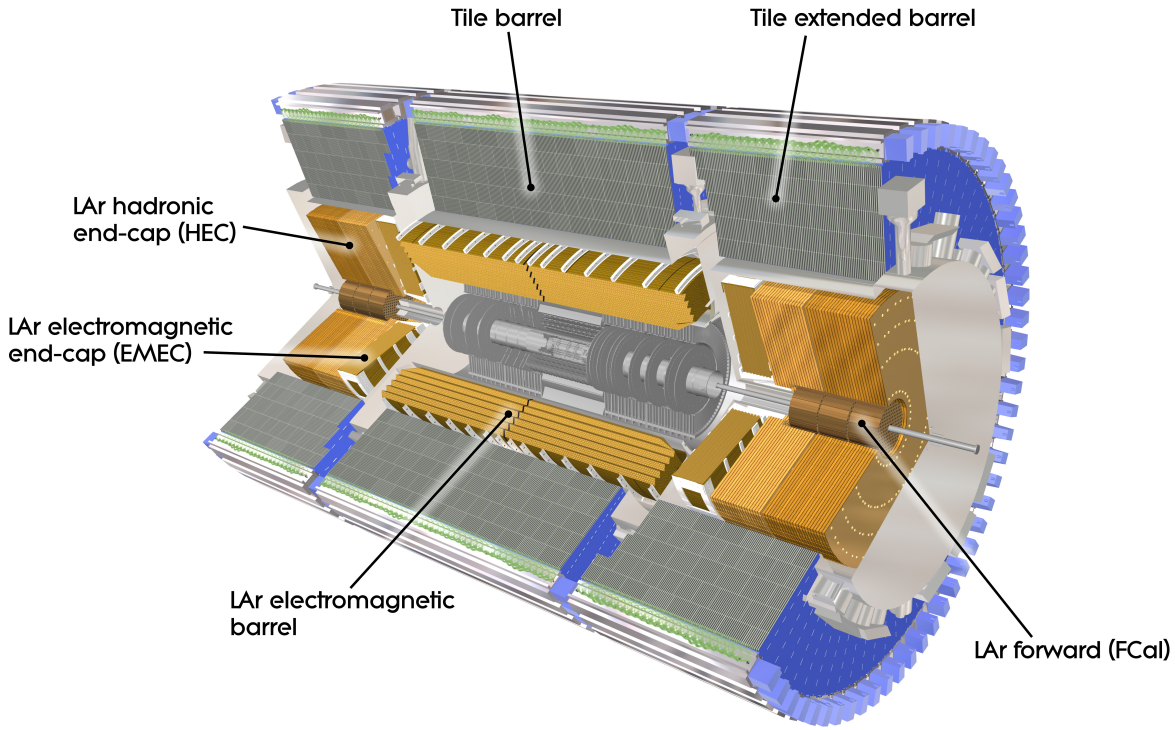


Figure 2.4: The calorimetry layout. Illustration from CERN Document Server.

via a top-loop, as well as detection of exotic particles, e.g. W' and Z', decaying to electrons. The ECAL is a lead-LAr detector, meaning that it essentially consists of a stack of lead plates immersed in liquid argon (LAr). Applying voltages between the lead plates creates a series of ionization chambers. The lead-plate absorbers and kapton-insulated copper electrodes are shaped as an accordion which provide a crack-less coverage in ϕ. In total, the ECAL covers $|\eta| < 3.2$ and has a total thickness of > 22 radiation lengths in the barrel region and > 24 in the end-caps.

A presampler detector in the region $|\eta| < 1.8|$ is used to correct for the energy lost in the ID, the solenoid magnet and the cryostat. It also contributes to the measurement of the EM shower directions and background rejection, e.g. π^0-background for photons and π^0-$\pi^\pm$ overlap as a background for electrons. The difficult transition region, $1.37 < |\eta| < 1.52$, between the barrel and end-cap cryostats is not used for photon identification nor precision measurements with electrons. This is due to a significantly degraded energy resolution despite the presence of the presampler.

Hadronic calorimetry

Although the HCAL has a coarser granularity than the ECAL, it is sufficient for meeting the requirements of jet reconstruction and E_T^{miss} measurements. It also enhances the particle identification of the ECAL by measuring e.g. leakage and isolation.

The barrel part is a tile-based sampling calorimeter divided into three cylinders covering the region $|\eta| < 1.7$. While scintillating tiles are the active material, steel acts as the absorber. When hadrons interact with the steel, more particles are created. This burst of particles excites the atoms of the scintillator causing them to emit photons, which in turn are collected by

wavelength-shifting fibres at the tile edges. The wavelength-shifted light can then be converted to electric signals by photomultipliers, giving a measurable electric pulse. The radial depth of the tile calorimeter is about 7.4 interaction lengths.

The HEC covers the range $1.5 < |\eta| < 3.2$, overlapping both the tile calorimeter and the FCal. This reduces the drop in material density in the transition areas. Due to the higher $|\eta|$, the HEC is more exposed to radiation. Parallel copper plates are thus chosen as absorber.

The last part of the HCAL, the FCal[4], provides coverage over $3.1 < |\eta| < 4.9$. Such a high $|\eta|$ and short distance (4.7 m) from the interaction point exposes it to high particle fluxes. Thus the LAr gaps are made as small[5] as 0.25 mm to avoid ion build-up problems. To achieve a high absorption length and limit the transverse spreading of showers, the modules FCal2 and FCal3 are mainly made of tungsten, while the FCal1 is made of copper, which optimizes both energy and position resolution.

2.2.5 Muon system

The muon system, see Figure 2.5, including the toroid magnets is the outermost sub-system and defines the overall dimensions of ATLAS. Muons are approximately 200 times heavier than electrons and are long-lived. Thus they can traverse the calorimeters without being stopped, and an additional system is needed to identify them and measure their momenta. Muons are deflected by the large air-core toroid magnets and their tracks measured by high-precision tracking chambers divided into four subsystems arranged in three layers. The muon spectrometer (MS) covers the region $\eta < 2.7$.

Precision measurements of track coordinates in the bending plane are performed by the Monitored Drift Tube (MDT) chambers which span most of the $|\eta|$-range. They consist of three to eight layers of drift tubes, giving a resolution of about 35 μm per chamber. In the innermost plane in the forward region $2 < |\eta| < 2.7$, Cathode Strip Chambers (CSCs) with higher granularity and higher rate capability are used. These are multiwire proportional chambers allowing two-dimensional measurements. Together with fine granularity, this makes it possible to resolve tracks in the forward region where track density is higher than in the barrel region. The resolution is 40 μm in the bending plane and 5 mm in the transverse plane. The disadvantage of the drift tubes is the relatively large drift time. Therefore a special layer of trigger chambers is needed.

The trigger system, which consists of Resistive Plate Chambers (RPCs) in the barrel region and Thin Gap Chambers (TGCs) in the end-cap regions, searches for patterns of hits consistent with high-p_T muons coming from the interaction region. Covering the range $|\eta| < 2.4$, it provides bunch-crossing identification, has well-defined p_T thresholds and measures the track coordinates in both the bending plane and the transverse plane. After matching hits in the bending plane of the MDT and the trigger chamber, the transverse coordinate of the trigger chamber is adopted as the second coordinate of the MDT measurement.

2.2.6 Trigger

With 40 million events per second and a raw data size of 25 MB per event, ATLAS produces an enormous flow of data. ATLAS's trigger system has three levels, L1, L2 and the event filter, where each level refines the analysis of the previous level. The proposed EF rate at a luminosity of 10^{33} cm^{-2}s^{-1} is 200 Hz, with an event size of approximately 1.3 MB. However, attempts will be made to keep the peak rate of the 2011 proton-proton running at 400 Hz.

[4]Actually, the FCal consists of three modules; the electromagnetic FCal1, and the two hadronic FCal2 and FCal3.

[5]The gaps of the ECAL barrel measure 2 mm.

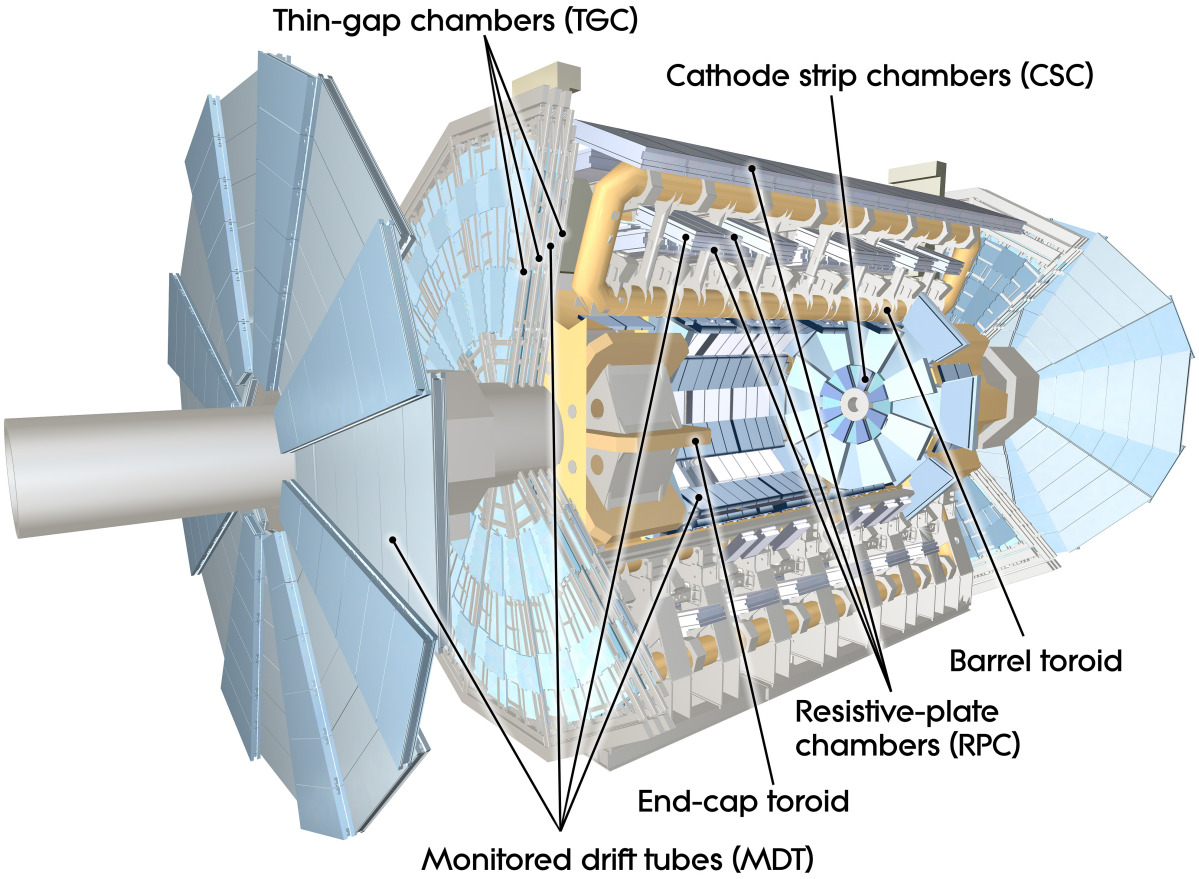


Figure 2.5: The muon system layout. Illustration from CERN Document Server.

The first trigger level, L1, only uses a limited amount of information to make a decision. It searches for high-p_T muons, electrons, photons, jets and taus decaying into hadrons, but also large missing and total E_T. So-called Regions-of-Interest (RoI's), regions whitin the detector where interesting features are identified, are also defined. At this level, only information from the calorimeters and the muon system are used. The L1 trigger makes a decision in less than 2.5 μs after the bunch-crossing and has an event rate of 75 kHz. The results are processed by the central trigger processor which implements adjustable selection algorithms.

Together, the L2 and the event filter form the High-Level Trigger (HLT). Full granularity and precision data of the calorimeter and muon system are employed, as well as data from the inner detector. The L2 trigger reduces the event ratio to about 3.5 kHz using full precision data from the RoI's. By applying offline analysis procedures, the event filter reduces the event rate even more, down to a rate of up to 200 Hz. The average event processing time of the L2 and event filter are 40 ms and 4 seconds, respectively. In Figure 2.6 the trigger and data acquisition systems are illustrated.

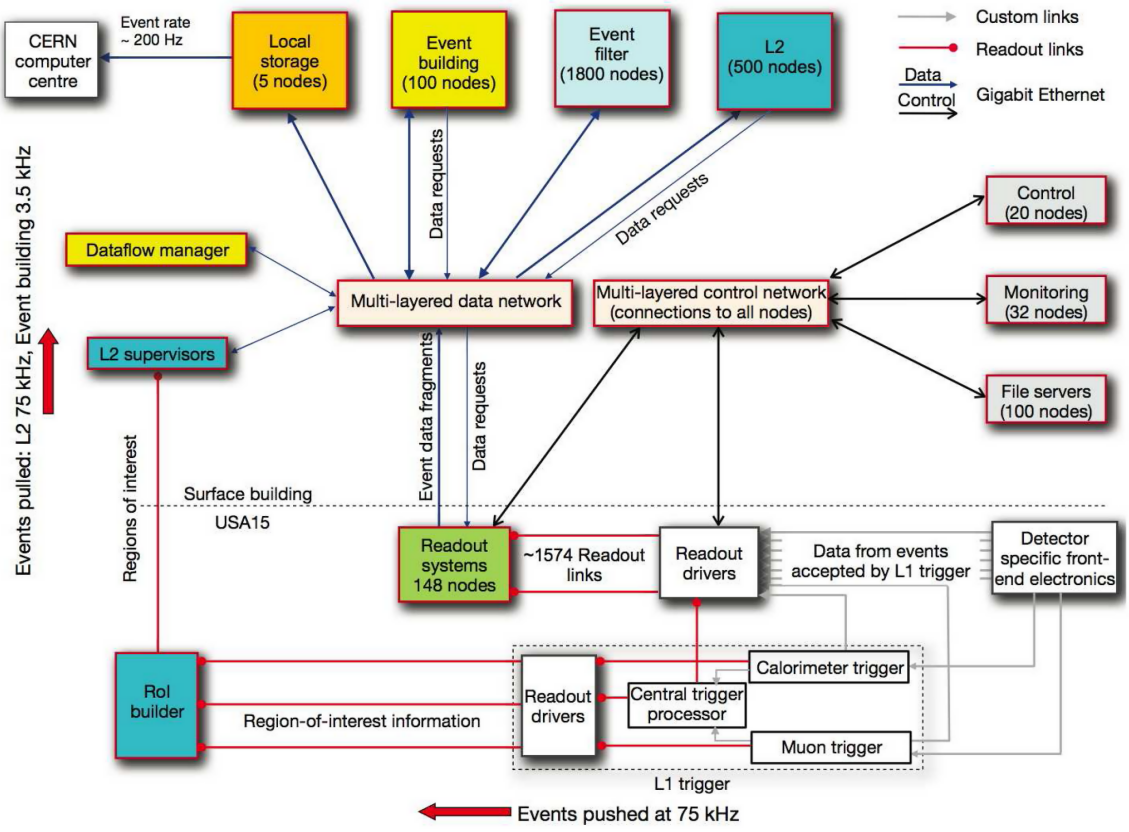


Figure 2.6: The ATLAS trigger and data acquisition systems. Illustration from [35].

Chapter 3

Monte Carlo, simulation and reconstruction

Simulation of physical processes is an essential part of any high energy physics experiment and serves as a reference when analyzing real data. As the nature of quantum physics is statistical, large samples of simulated data are needed. These are generated by means of *Monte Carlo* (MC) techniques, which simulate assumed probability distributions. In general, numerical integration methods approximate the integral

$$\int_{x_{min}}^{x_{max}} f(x)dx \approx \sum_{i=1}^{N} w_i f(x_i), \tag{3.1}$$

where w_i are the weights for each discrete value of x. Compared to other methods, the Monte Carlo method is highly efficient and is thus widely used in physical sciences.

There are many Monte Carlo techniques, but the possibly most common situation is that a function $f(x)$, which is non-negative in the defined range of x, is known. For each randomly selected x the probability in a small interval dx around x is proportional to $f(x)dx$. If $f(x)$ is a differential cross section, the Monte Carlo integration will then give an estimate of the cross section, but it can be any other distribution as well. By sampling N random x from a given sampling interval, the expected value of f in this interval is given by

$$\langle f \rangle = \frac{1}{N} \sum_{i=1}^{N} f(x_i)p(x_i), \tag{3.2}$$

where $p(x)$ is an assumed distribution. The task of a Monte Carlo-based event-generation is both to generate events and to estimate the total cross section.

However, the generated events are not directly comparable to data. In order to do so, they must undergo what can be thought of as a series of *transformations*. This involves simulating the response of the various detectors and triggers and processing the information with reconstruction software. What follows is a description of how the full simulation chain is brought about in the ATLAS software. With this thesis being the first study of black holes in di-lepton final-states in ATLAS, several PBH samples were generated, simulated and reconstructed for private use. The production procedure will be explained, and branching ratios and relevant efficiencies will be shown.

3.1 The full simulation chain

In the ATLAS offline software [36], going from Feynman rules to tangible simulated data is done in three main steps. First, the physics events are generated by event generators, producing an *event record*. This is given as input for simulation, the second step, in which the geometry of ATLAS, the material and how particles interact with it is simulated. The output is a collection of *hits*, energy deposits in sensitive detector volumes, and these are further processed by the digitization algorithm. At this stage, the hits are translated to the output actually produced by the electronic readout of the different detectors. From raw data, the reconstruction algorithm collects all information and derives particle parameters.

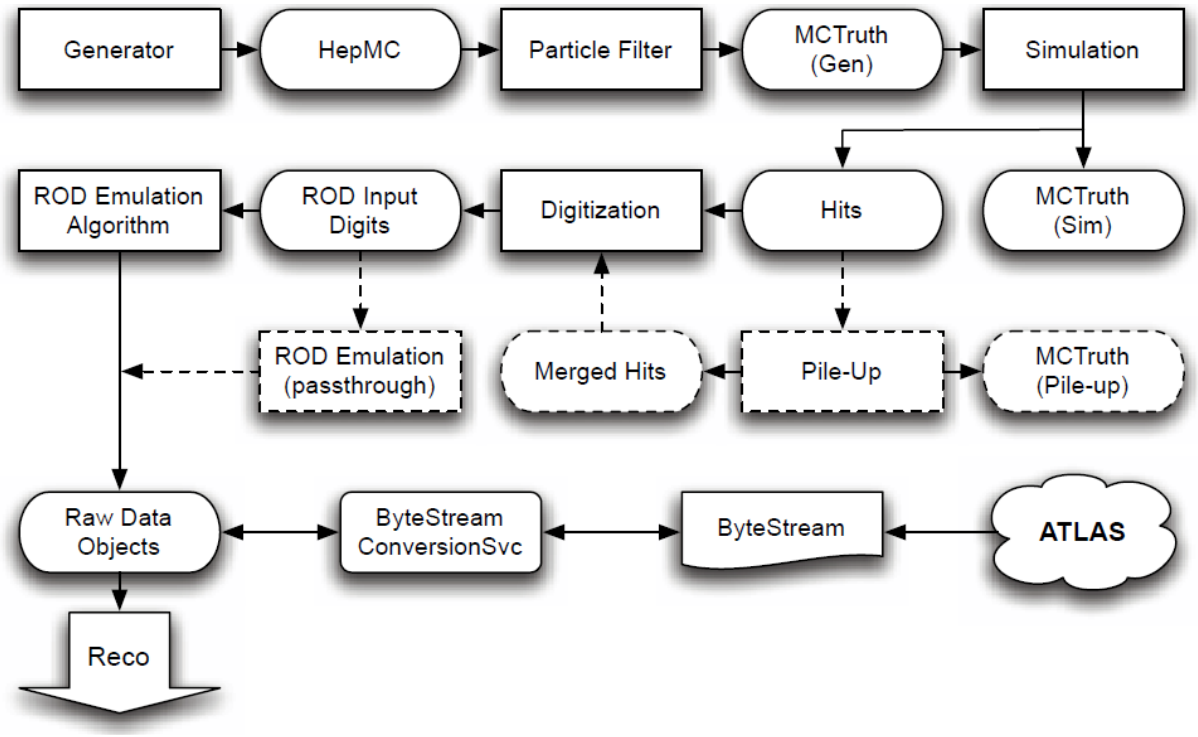


Figure 3.1: The data flow of the simulation step. The round-cornered boxes represent objects of data, while the square-cornered represent processing stages. Diagram from [36].

3.1.1 Event generation

The event generators model the physics of hard processes, radiation in the initial or final state, beam remnants, hadronization and prompt decays of unstable particles, such as Z and W bosons. While the general-purpose generators usually handle most of the mentioned effects, parton-level generators only model a specific particle scattering process, e.g. $pp \rightarrow Z^0 \rightarrow \mu^+\mu^-$. The signal samples used in this analysis were generated by the parton-level Monte Carlo generator BlackMax [37], interfaced to PYTHIA [38], a general-purpose Monte Carlo generator, for hadronization, prompt decays and underlying event modeling. All stable particles expected to propagate through some part of the detector are stored in the event record.

BlackMax is a comprehensive BH-event generator accounting for several BH models, including the two-particle final-state scenario described earlier. While BlackMax only accounts for the direct decay of the BH, Pythia models prompt decays and hadronization of the two outgoing particles. Pythia is run from within the Athena framework [39][40], where the user may apply filters to reject uninteresting events. The output, converted into a common format by mapping

into the object-oriented MC event record HepMC [41], contains the so-called MCTruth information where the true properties of the particles are stored. No interaction with the matter content of the detector, external fields, noise in readout channels and other effects are considered at this stage.

3.1.2 Simulation

In the ATLAS offline software, full simulation is done with Geant4 [42], while fast simulation is done with Atlfast [43]. As input, the collection of four-vectors corresponding to the stable particles stored in the event record is given. Geant4 has a rich functionality including solutions for geometry and material description, propagation of particles through the detector, modelling of physics processes and the concept of sensitive detectors. The latter allows for defining active detector elements, which trigger the recording of hits when particles interact with them. Information such as position, energy deposit, identifier of the active element and so on is carried by the hits.

Atlfast-II gives an output identical to standard simulation and reconstruction, with exactly the same naming of objects as in full simulation. It provides a full simulation of the inner detector, while the calorimeter is simulated using the FastCaloSim [44] algorithm (see section 3.3.1) based on per-particle average shower-shapes. It is also possible to reduce the running time further by replacing the ID simulation with a parametrized distribution of track hits.

After simulation, the hits are converted to detector responses, which are typically voltages or times on pre-amplifier outputs. As well as accounting for the propagation of charges or light in the detector parts, a highly detector-specific task, electronics noise and channel-dependent variations in detector response are modelled. The final output are Raw Data Objects (RDOs) resembling real detector data, e.g. voltages, times and positions.

3.1.3 Reconstruction

From RDOs, the reconstruction algorithm derives information necessary for most physics analysis and produces Event Summary Data (ESD). This includes reconstruction of the tracking and calorimetry detectors, sufficient for identification of particles, jet calibration and track fitting. Combining information from all detectors, objects such as photons, electrons, muons, taus, jets, missing transverse energy E_T^{miss} and primary vertices are built. Furthermore, the reconstruction of complex objects, such as the b-tagging object, is stored as Analysis Object Data (AOD) [45]. The reconstruction of electrons and muons will be discussed in more detail in Section 3.4.

Datasets used in this thesis are Derived Physics Datasets (DPDs), more specifically D3PDs. A DPD is a reduced version of the full dataset (ESD, AOD), where the level of detail may vary. Very detailed information not needed for the analysis, e.g. error matrices from tracks or calorimeter cells which are not in the vicinity of interesting physics objects, are removed. The analysis framework used is ROOT [46], and D3PDs are simply DPDs in a ROOT ntuple format.

3.1.4 Job transforms

The so-called *job transforms* are python scripts used to run production tasks. They transform the input file into one or more output files in a format recognized by the ATLAS offline software, the POOL format [48]. Depending on the transform script, there are a number of required and optional parameters. A transform script with a specific set of parameters corresponds to an *ATLAS Metadata Interface (AMI) tag*.

Sample name	Mth[M_P]_c2_n6	Mth[M_P]_c1_n6	Mth[M_P]_c3_n6
Incoming particles	pp	pp	pp
Center of mass energy	7000 GeV	7000 GeV	7000 GeV
Planck scale in D dimensions	1500/1750/2000 GeV	1500/1750/2000 GeV	1500/1750/2000 GeV
Planck scale convention	1 (default, Eq. 1.27)	1	1
Case	4 (BH→ 2)	4 (BH→ 2)	4 (BH→ 2)
Number of extra dimensions	6	6	6
PDF	CTEQ6.6 [47]	CTEQ6.6	CTEQ6.6
Minimum mass	M_P	M_P	M_P
Maximum mass	7000 GeV	7000 GeV	7000 GeV
Definition of cross section	2 (Eq. 1.36)	2	2
Calculate cross section according to	1 (center-of-mass energy)	1	1
Loss factors	0	0	0
Graviton	No	No	No
Conservation laws	B - L, B, L, flavours (9)	B - L, B, L	B - L

Table 3.1: BlackMax parameters used in the sample production.

There are transform scripts for each of the before-mentioned simulation steps, including scripts which combine two or more steps. Some parameters, such as the geometry [49] and conditions [50] of the ATLAS detector, must be consistent throughout the chain of transforms. To avoid any problems, it is usually a good idea to use the options corresponding to the AMI-tags of an officially produced dataset. Preferably, to ensure consistency, i.e. identical conditions and tools applied, all datasets used in an analysis should have the same, or at least similar, tags.

3.2 Event generation with BlackMax

BlackMax simulates semi-classical and Planckian black hole production and evolution in the context of brane world models with low-scale quantum gravity. Being based on phenomenologically realistic models, the generator avoids the serious problems plaguing low-scale gravity, e.g. fast proton decay. In the split-fermion model [51], for example, the SM fields are confined to a brane much thicker than M_P^{-1}. Quarks and leptons, being stuck on each three-dimensional slice, are separated by much more than M_P^{-1}, causing any direct coupling to be exponentially suppressed. This ensures proton longevity.

In BlackMax, the two-particle final state scenario models non-rotating black holes on a tension-less brane. As is described more in detail in the documentation [37], there are several parameters affecting BH formation and decay. Fermion splitting, rotation, brane tension and recoil due to Hawking radiation are just some of them, and so the two-particle final state scenario, without any of these effects, is a naive and minimalistic, but general model.

The user is able to choose among different models, define conservation laws and otherwise set the relevant parameters for the model chosen [52]. The samples produced for this thesis will use the same parameters as an official dataset (run number 105453) and these are listed in Table 3.1. Parameters omitted from the table are either not applicable to the two-particle final state scenario, or the default is used. The parton PDF interface used is LHAPDF [53].

The BH initial state is in BlackMax characterized by

$$\begin{aligned} E &= E_{in} f_E; \\ P_z &= P_{in} f_P; \\ J' &= L_{in} f_L; \end{aligned} \tag{3.3}$$

where E_{in}, P_{in} and L_{in} are initial energy, momentum and angular momentum of the colliding partons, while f_E, f_P and f_L are the fractions of these retained by the stationary BH. It should be noted that the recommended values for the loss factors, $1 - f_E$, $1 - f_P$ and $1 - f_L$ are 10% to 15%. However, the official dataset and the datasets for this thesis were produced with all set to zero, the major consequence being that the calculated cross sections are larger than the true ones.

3.2.1 Branching ratios

Leptonic decays are much rarer than hadronic in the two-particle final-state scenario with proton-proton collisions. In order to get a reasonable amount of statistics for the leptonic decays, it is necessary to generate a huge amount of BH events. As the simulation and reconstruction process is quite time-consuming, an event selection was implemented already at the parton-level generation step. This was easiest done by modifying BlackMax's algorithm which writes events to the event record. In order to obtain the desired number of events for the most relevant decay channels, they were classified and written to separate event records. To get the cross section for each class of decay channels, the total cross section is multiplied with the relevant branching ratio.

M_P (GeV)	1500			1750			2000		
Total cross section (pb)	994 ± 5.46			344 ± 1.99			125 ± 7.60		
N conservation laws	1	3	12	1	3	12	1	3	12
Branching ratio (%)									
e^+e^-	0.025	0.025	0.037	0.021	0.021	0.032	0.018	0.018	0.027
$\mu^+\mu^-$	0.024	0.025	0.037	0.021	0.021	0.031	0.017	0.018	0.027
$e^\pm + \mu^\mp$	0.048	0.049	0	0.042	0.042	0	0.036	0.036	0
$\gamma\gamma$	0.012	0.013	0.019	0.011	0.010	0.015	0.009	0.009	0.013
$Z + X$	3.13	3.14	3.80	2.74	2.73	3.31	2.41	2.41	2.92
$2 \times \{e, \mu, \tau, W, b, t\}$	11.24	8.50	0.94	11.93	8.82	0.78	12.53	9.09	0.66
$X + X$	85.52	88.26	95.17	85.23	88.36	95.83	84.98	88.42	96.35

Table 3.2: PBH decay channels with branching ratios for different Planck masses and conservation laws. The ratios are based on $> 10^7$ generated events for each point in parameter space.

In Table 3.2 the decay channel classes along with total cross sections and branching ratios are listed. One can see that the leptonic decay channels have a mere branching ratio of ∼0.05 - 0.1%, but there are other channels which may give a clean lepton signal as well, e.g. $PBH \to Z + X$. From earlier experiments it is known that the $Z \to ee/\mu\mu$ branching ratio is ∼6.7%, such that ∼0.2% of the PBHs decay to di-leptons via a Z boson.

In case of confusion, the di-lepton and $2\times\{e, \mu, \tau, W, b, t\}$ samples do not overlap. A di-lepton event is always and only written to its respective event record. Hence, the $2 \times \{e, \mu, \tau, W, b, t\}$ decay channels are in the case of flavour-conservation reduced to $2 \times \{\tau, W, b, t\}$. When turning off conservation of flavour, the $2 \times \{e, \mu, \tau, W, b, t\}$ branching ratio increases drastically. This can simply be explained by the parton PDFs. A very small fraction of the sea quarks within the proton are heavy flavours, i.e. flavour-conserving PBHs seldom decay to bottom or top. In addition, electron + tau and muon + tau final states are possible. As soon as baryon and lepton number conservation are turned off, final states such as lepton + jet are made possible, and the branching ratio of the $2 \times \{e, \mu, \tau, W, b, t\}$ channels increases even more.

In order for a PBH to decay into a pair of leptons it must be colour neutral, and thus the incoming partons must be either two gluons or a quark and an anti-quark. As the LHC is colliding protons, the anti-quark PDFs are softer[1] than the quark PDFs, and so are the gluon PDFs. This is clearly seen in Figure 3.2 and results in di-lepton branching ratios which decrease as the Planck mass, and thus the PBH threshold, increases.

Some of the generated samples contain mostly uninteresting events, and these can be filtered out before starting the simulation by applying a two-lepton filter. If there are at least two leptons in the event after unstable particle decay, the event is kept. The filter efficiency is then defined as

$$\varepsilon_C = N(C)_{LL}/N(C), \tag{3.4}$$

where C is the class of decay channels. In Table 3.3 the two-lepton filter efficiencies obtained for the relevant samples are listed. One might expect that the filter efficiency for $Z + X$ would be close to the branching ratio of leptonic Z decay, but it is significantly larger. The Z can decay to other unstable particles, and both these and the other particle X can decay to leptons.

[1]Here soft refers to a small x, where x is the fraction of the proton energy carried by the parton. The distributions are functions of the momentum transfer, Q^2, between the two colliding partons.

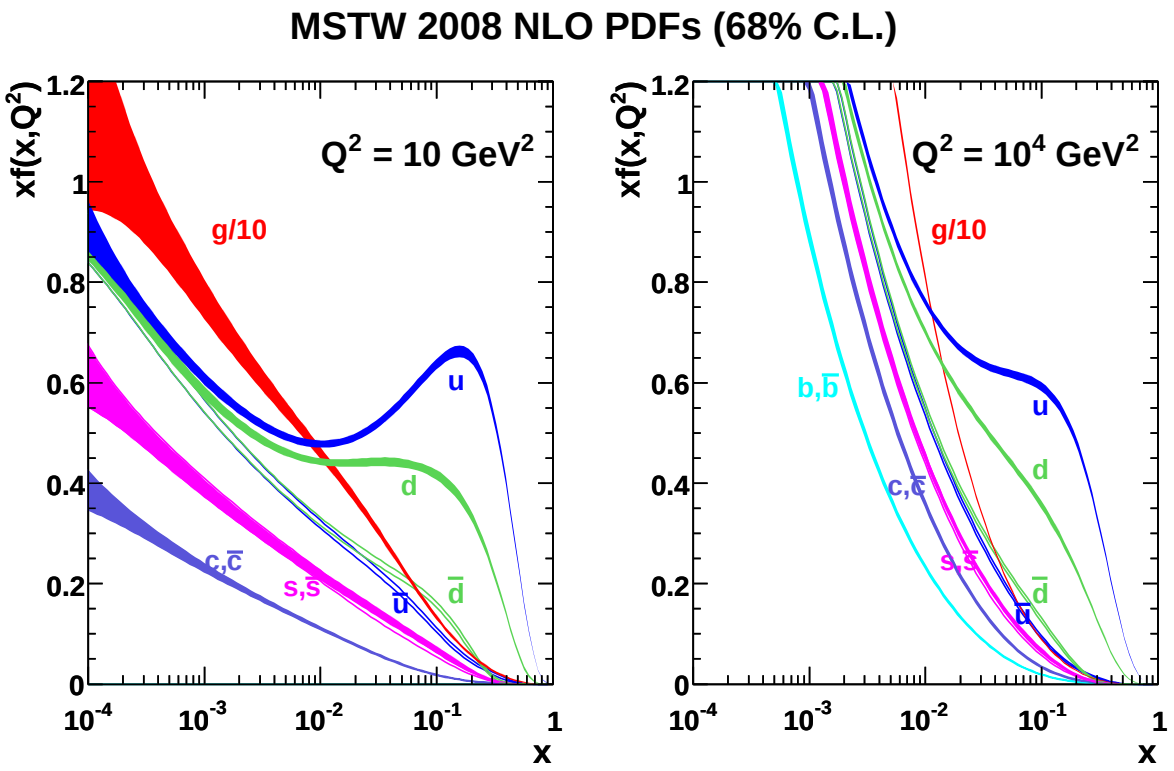


Figure 3.2: The Martin-Stirling-Thorne-Watt parton distribution functions for 10 GeV (to the left) and 10 TeV (to the right) momentum transfer. Plot from http://projects.hepforge.org/mstwpdf/.

M_P (GeV)	1500			1750			2000		
N conservation laws	1	3	12	1	3	12	1	3	12
Two-lepton filter efficiency, ε_C (%)									
$Z + X$	13.7	13.6	10.5	14.5	14.3	10.9	15.1	15.1	11.3
$2 \times \{e, \mu, \tau, W, b, t\}$	26.9	26.8	23.8	28.0	29.1	25.2	29.3	30.7	25.9
$X + X$	6.04	6.21	2.01	6.60	6.95	2.18	7.39	7.54	2.49

Table 3.3: Two-lepton filter efficiency.

3.2.2 Defining conservation laws

The desired conservation laws can be specified in the parameter file in the following way:

```
number_of_conservation
1
d,s,b,u,c,t,e,mu,tau,nu_e,nu_mu,nu_tau
1,1,1,1,1,1,0,0,0,0,0,0
```

The *number_of_conservation* parameter specifies the number of conservation rules subsequent to the *d,s,b,u,c,t,e,mu,tau,nu_e,nu_mu,nu_tau* line. Each rule is then specified as an ordered set of integer coefficients a_f such that $\sum_f a_f N_f$ is a constant, with N_f being the number of the fermion flavours. The rule in the given example yields baryon number conservation. This follows from

$$\sum_f a_f N_f = N_u + N_d + N_c + N_s + N_t + N_b = k, \tag{3.5}$$

where the leptons play no role in baryon number conservation, hence the zeros. Similarly, if one in addition wishes to conserve the lepton number, another line has to be added:

```
number_of_conservation
2
d,s,b,u,c,t,e,mu,tau,nu_e,nu_mu,nu_tau
1,1,1,1,1,1,0,0,0,0,0,0
0,0,0,0,0,0,1,1,1,1,1,1
```

Conservation of B - L yields

$$\frac{1}{3}\sum_f N_q - \sum_f N_l = k, \tag{3.6}$$

which can be specified in the following manner:

```
number_of_conservation
1
d,s,b,u,c,t,e,mu,tau,nu_e,nu_mu,nu_tau
1,1,1,1,1,1,-3,-3,-3,-3,-3,-3
```

It is also possible to conserve all fermion flavours individually. In this case, nine lines must be added:

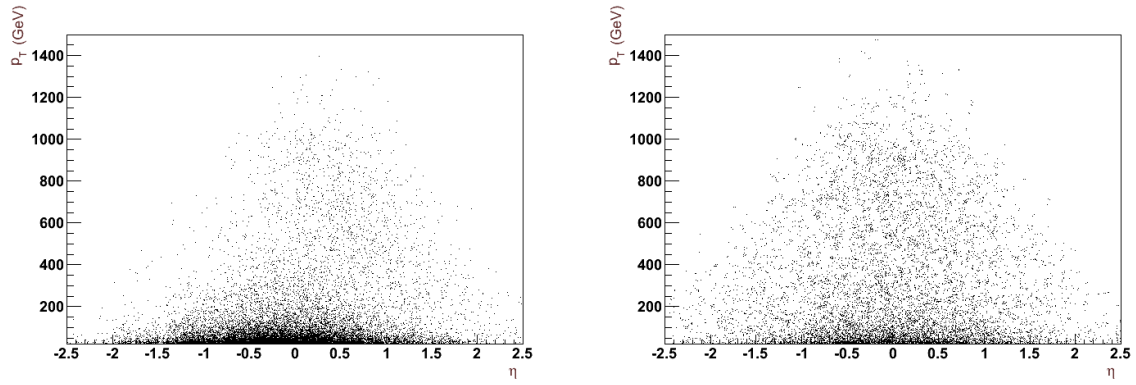
```
number_of_conservation
9
d,s,b,u,c,t,e,mu,tau,nu_e,nu_mu,nu_tau
1,0,0,0,0,0,0,0,0,0,0,0
0,1,0,0,0,0,0,0,0,0,0,0
0,0,1,0,0,0,0,0,0,0,0,0
0,0,0,1,0,0,0,0,0,0,0,0
0,0,0,0,1,0,0,0,0,0,0,0
0,0,0,0,0,1,0,0,0,0,0,0
0,0,0,0,0,0,1,0,0,1,0,0
0,0,0,0,0,0,0,1,0,0,1,0
0,0,0,0,0,0,0,0,1,0,0,1
```

The number of conservation laws does not affect the production cross section, but alters the branching ratios as seen in Section 3.2.1.

3.2.3 Bug-fix

Before starting a full production, small testing samples were generated with BlackMax 2.01.3. It soon became clear that there was an asymmetry in the pseudorapidity of the decay products.

Using pure generator level information, the electrons from a PBH Z showed a tendency to have positive η values, meaning that the Z in some way preferred to be emitted in the positive η direction. There is no physics explaining such a phenomena, so it had to be a bug. The authors were contacted, and as the bug was already known, they provided me an unreleased bug-fix version (REV.62). This version is very similar to 2.02.0 which was later released, but the latter also dealt with some of the memory leaks in earlier releases. These may pose a problem when generating a large number of events. In Figure 3.3 a comparison between version 2.01.3 and 2.01.3 REV.62 is shown, and one can see a great improvement in the latter.



(a) Version 2.01.3

(b) Version 2.01.3 REV.62

Figure 3.3: p_T (GeV) distribution w.r.t. η of all generator level electrons in $PBH \rightarrow Z + X$ samples generated with v. 2.01.3 (a) and v. 2.01.3 REV.62 (b). The difference in number of entries is just due to the sample sizes not being equal.

3.3 Atlfast-II simulation

All PBH samples for the feasibility study in the next chapter were produced with the Atlfast-II-D (default) algorithm, a combination of fast and full simulation. The advantage is a significant reduction in CPU time by using parametrizations of the ECAL and HCAL showers. Just as for full simulation, production of II-D samples runs in three steps, simulation, digitization and reconstruction. In the simulation step, a normal Geant4 full simulation is used for the inner detector, in which all particles but muons are killed at the exit of the detector volume. The calorimeters are simulated using FastCaloSim, which runs in the digitization step. Its output is converted into Geant4 hits and further processed by the standard calorimeter digitization. Muons are however fully simulated in all systems, ID, calorimeters and MS. Simulation of trigger information is run in the normal way as well.

In producing the PBH samples, the transform options for ATLAS offline software Release 15.6.10.6 based on the full simulation tags s838 and r1307 were used [54]. The options given for each transform are listed in Appendix D.1.

3.3.1 Calorimeter simulation model

The detailed Geant4 simulations require a CPU time of several minutes per event, and more than 90% of this time is spent inside the calorimeter systems. This is a challenge for the production of sufficiently large MC samples. The FastCaloSim package provides an accurate but fast detector simulation, a nice compromise. By parametrizing the response and energy distributions in the

ATLAS detector, the calorimeter simulation is reduced to only a few seconds per event. In order to achieve such improvement in speed, some simplifications were made.

- Calorimeter cells are described as cuboids in η, ϕ and ρ. In the homogeneous regions of the ECAL this is a reasonable assumption. For the other calorimeters and the edge regions, it is only an approximation.
- Particle showering simulation is replaced by parametrizations. The longitudinal shower properties with fluctuations and correlations are reproduced, but lateral shower properties and uncorrelated lateral energy fluctuations are only averaged. Although not adequate for hadrons, the ansatz of average lateral shape is well suited for photons and electrons. An improved model for hadrons is currently under development.
- There are three types of particles which are parametrized and used for the simulation: photons, electrons and charged pions. The latter is used for all hadrons, both neutral and charged.

3.3.2 Electron performance

In [55] the electron reconstruction and trigger efficiencies were compared for fast and full simulation. A sample of $Z \rightarrow e^+e^-$ events was generated with Pythia and then sent through both full Geant4 simulation and Atlfast-II. The same analysis code with no corrections was applied to both reconstructed samples. No p_T cut was placed, but the electrons were preselected within the pseudorapidity range $|\eta| < 2.47$, but not in the crack $1.37 < |\eta| < 1.52$ where the energy resolution is degraded. Figures 3.4(a) and 3.4(b) with captions are copied with permission from [44].

In Figure 3.4(a) the reconstruction efficiency for electrons is shown. The electrons were required to match a truth (generator level) electron within a distance $\Delta R < 0.1$ and pass medium quality cuts[2]. Truth electrons with $p_T > 5$ GeV and $|\eta| < 2.5$ were used to calculate the selection efficiencies. It was found that the Atlfast-II simulation reproduces the full Geant4 simulation within the statistical precision of 5% or better.

Figure 3.4(b) shows the combined trigger efficiency for the event filter trigger EF_e10_loose and the relevant Level 1 and Level 2 triggers. According to its name, the EF trigger selects loose quality electrons with an E_T threshold of 10 GeV. The trigger efficiency was calculated w.r.t. offline electrons passing the preselection. Even though the trigger fired, one doesn't know on which electron. To determine whether the selected offline electron was triggered or not, it was required to be matched within a distance $\Delta R < 0.2$ to a relevant trigger Region-of-Interest. The trigger efficiency for the Atlfast-II simulation was found to reproduce the Geant4 simulation within the statistical precision of 5% or better.

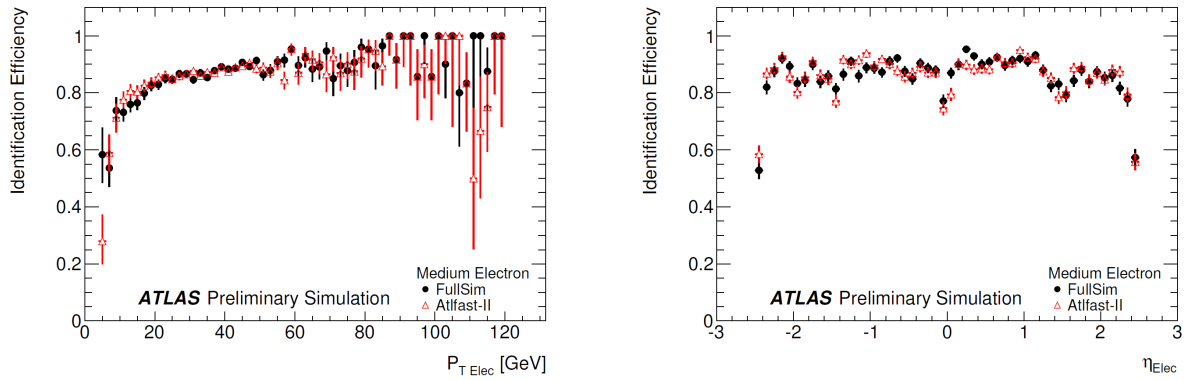
3.4 Reconstruction of leptons

Before proceeding it is essential to discuss the reconstruction and identification of electrons and muons in more detail. At the end of this section, the direct lepton-decay PBH samples produced with Atlfast-II will be examined.

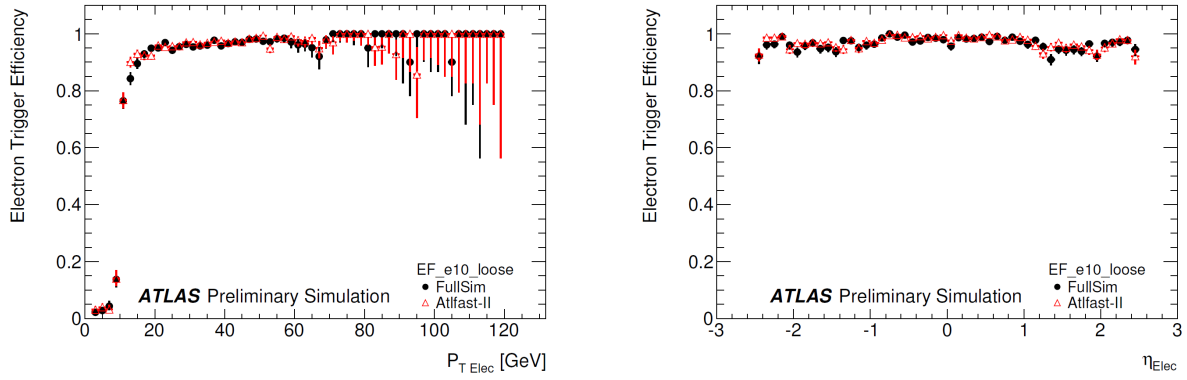
3.4.1 Electrons

The standard (cluster-based) electron and photon reconstruction and identification algorithm [56] starts from a calorimeter seed and is optimized for high identification efficiency for calorime-

[2]The different levels of electron quality cuts are discussed in Section 3.4



(a) Electron identification efficiency for medium quality electrons, with respect to truth electrons, as a function of pT (left) and pseudorapidity (right). The black points show the full Geant 4 simulation and the red points the Atlfast-II simulation.



(b) Trigger efficiency of the EF_e10_loose trigger with respect to offline loose electrons, as a function of pT (left) and pseudorapidity (right). The black points show the full Geant 4 simulation and the red points the Atlfast-II simulation.

ter transverse energies greater than 20 GeV. If a seed electromagnetic cluster with E_T above 2.5 GeV is found in the second layer of the ECAL, a matching track is searched for among all reconstructed tracks not belonging to a photon-conversion electron reconstructed in the ID. After extrapolating the track to the ECAL, it is required to match the cluster within a broad window of $\Delta\eta \times \Delta\phi = 0.05 \times 0.10$. The ratio of the energy of the cluster to the momentum of the track is required to be < 10. For true isolated electrons with $E_T > 20$ GeV and $|\eta| < 2.5$, these requirements result in an identification effiency of 93%. There are other algorithms as well, e.g. the soft (track-based) algorithm and the "forward" electron reconstruction algorithm.

Further quality requirements may be applied in the end-user analysis. There are three reference sets of cuts which have been defined as *loose*, *medium* and *tight*. In high-mass di-electron final states the electron p_T typically ranges from 100 GeV up to several TeV. The background for very high-p_T electron is expected to be small, such that loose or medium cuts are sufficient, retaining a high efficiency.

The loose cuts are based on limited information from the calorimeters and are applied on hadronic leakage and shower-shape variables. A very loose matching between reconstructed track and calorimeter cluster is required. This simple set of cuts gives excellent identification efficiency while maintaining rejection against highly energetic pions with wide showers.

The medium cuts have additional cuts on the strips in the first layer of the ECAL and on the tracking variables, e.g. number of pixel hits. The identification efficiency drops by $\sim 10\%$, but increases the jet rejection by a factor of 3-4. By exploiting the very fine granularity of the innermost layer of the ECAL and demanding higher track quality, neutral pion rejection increases.

The tight cuts includes cuts on the number of hits in the innermost Pixel layer, to further reject photon-conversion electrons, and in the TRT. To reject charged hadrons, a cut on the ratio of high-threshold hits to the number of TRT hits is placed. A proper matching between position and momentum in the cluster and extrapolated track is also required.

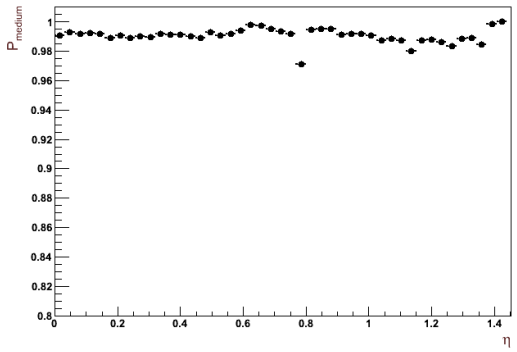
New electron quality definitions

It was found that data simulated with Release 15 of the ATLAS offline software mis-modeled the R_η and W_η distributions. These are both based on shower measurements in the second sampling layer of the ECAL. While W_η is a measure of the shower width, R_η is the ratio in η of cell energies in 3×7 versus 7×7 cells. Due to the mis-modeling, the loose, medium and tight definitions were loosened w.r.t. R_η and W_η. The new definitions are RobustLoose, RobustMedium and RobusterTight, and will in this thesis be used when analyzing Release 15 simulated data.

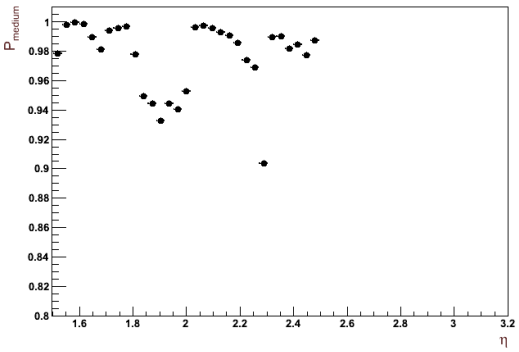
A $Z \to e^+e^-$ Monte Carlo dataset (run number 106046) was used to find the probability for a RobustMedium electron satisfying $p_T > 20$ GeV and $|\eta| < 2.47$ to pass the standard medium cuts as well. This was found to be 98.6%. In Figure 3.4 the probability for a RobustMedium electron to pass medium has been plotted as a function of $|\eta|$. Several structures can be seen, and a possible explanation for these are sudden falls or rises in the material distribution in the ID and ECAL which have been badly accounted for in the medium cuts. The radiation length in front of and in the ECAL is shown in Figure 3.5, whereas Figure 3.6 show the radiation and interaction length at the exit of the ID envelope. The figures with captions are taken from [35].

3.4.2 Muons

There are several ways of identifying and reconstructing muons. *Standalone* muons are only associated with a track in the muon spectrometer which is extrapolated to the beam line. *Combined* muons are found by matching standalone muons to tracks in the ID. Combining MS



(c) w.r.t. η



(d) w.r.t. p_T

Figure 3.4: The probability for a RobustMedium electron to pass medium as a function of $|\eta|$ in barrel (a) and end-cap (b).

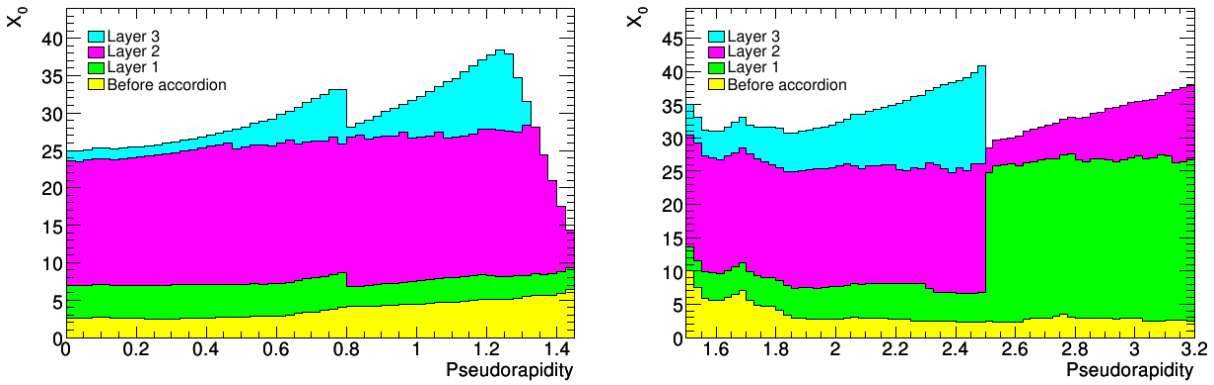


Figure 3.5: Cumulative amounts of material, in units of radiation length X_0 and as a function of $|\eta|$, in front of and in the electromagnetic calorimeters. The figures show, separately for the barrel (left) and end-cap (right), the thickness of each accordion layer as well as the amount of material in front of the accordion. Plot and caption from [35].

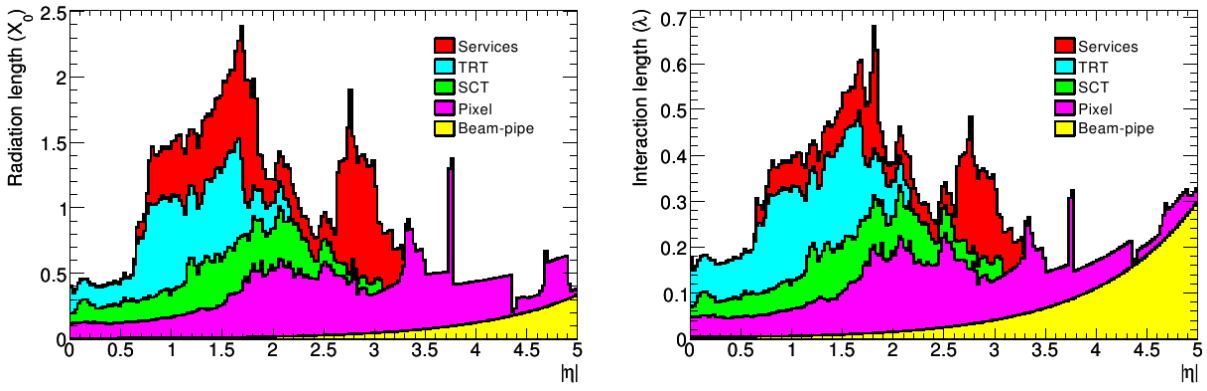


Figure 3.6: Material distribution (X_0, λ) at the exit of the ID envelope, including the services and thermal enclosures. The distribution is shown as a function of $|\eta|$ and averaged over ϕ. The breakdown indicates the contributions of external services and of individual sub-detectors, including services in their active volume. Plot and caption from [35].

measurements with the more precise pixel, SCT and TRT measurements, the bending of the muon within both the toroid and solenoid magnets is exploited. This increases the momentum resolution. By extrapolating tracks in the ID to the MS and searching for nearby hits, so-called *tagged* muons are found.

The MS has a slighly greater $|\eta|$ coverage than the ID, 2.7 and 2.5, respectively, an advantage for standalone algorithms. However, muons produced in the calorimeter from e.g. meson decay, serving as an additional background of fake muons, are often found in standalone reconstruction. Thus, only combined muons will be used in this thesis.

The association between MS and ID tracks for combined muons is performed with a χ^2 test. This is defined from the differences between the respective track parameters weighted by their combined covariance matrix,

$$\chi^2 = (T_{MS} - T_{ID})^T (C_{ID} + C_{MS})^{-1} (T_{MS} - T_{ID}), \tag{3.7}$$

where T denotes a vector of five track parameters, expressed at the point of closest approach to the beam line, and C is its covariance matrix. The STACO algorithm calculates a statistical combination of the ID and MS track vectors to obtain the combined track vector,

$$T = \left(C_{ID}^{-1} + C_{MS}^{-1}\right)^{-1} \left(C_{ID}^{-1} T_{ID} + C_{MS}^{-1} T_{MS}\right), \tag{3.8}$$

while the Muid algorithm does a partial refit. It starts from the ID track vector and covariance matrix and adds the measurements from the MS track.

3.4.3 Atlfast-II reconstructed signal samples

In Section 3.3.2 a good agreement between fast and full simulation w.r.t. electron identification efficiency was shown. The reconstruction efficiency for the PBH samples produced with Atlfast-II can be evaluated by doing a simple analysis on the direct lepton-decay samples. If an event of the e^+e^- samples contains at least two electrons passing the object requirements listed below, it is accepted as a properly reconstructed event, and the same goes for the $\mu^+\mu^-$ samples. Events of the $e\mu$ samples must contain at least one lepton of each flavour satisfying the requirements.

Electrons

- Electron reconstructed by cluster-based algorithm (see Section 3.4.1)
- Cluster $p_T > 25$ GeV
- Cluster $|\eta| < 2.47$
- RobustLoose (see Section 3.4.1)

Muons

- Combined muon (see Section 3.4.2)
- Combined $p_T > 25$ GeV
- Combined $|\eta| < 2.4$

Following this approach, the reconstruction efficiencies obtained for all direct lepton-decay PBH samples are presented in Table 3.4. The efficiencies are consistent with what one can expect from fully simulated samples.

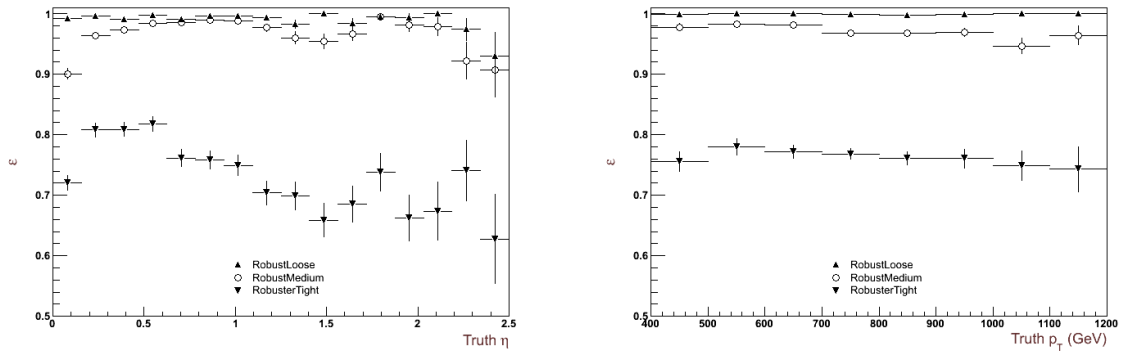
From Figure 3.7 it can be seen that the reconstruction efficiency is very high for both RobustLoose and RobustMedium electrons, whereas for RobusterTight it drops to $\sim 75 - 80\%$ in the barrel and $\sim 70\%$ in the end-caps and crack regions[3]. Not considering electron candidates

[3]The crack regions are $\eta = 0$, where the two barrel halves meet, and $1.37 > |\eta| > 1.52$, the region between the barrel and end-cap cryostats. The region between the barrel and end-cap cryostats suffers from significantly degraded energy resolution and is thus not used for precision measurements with electrons.

with $p_T > 400$ GeV, as these are mostly end-cap electrons, the reconstruction efficiency seems to be stable w.r.t. p_T.

M_P (GeV)	1500			1750			2000		
N conservation laws	1	3	12	1	3	12	1	3	12
Reconstruction efficiency (%)									
e^+e^-	91.1	91.2	91.1	90.5	91.3	90.4	91.5	91.9	90.7
$\mu^+\mu^-$	77.6	78.1	77.8	77.6	78.3	77.2	77.0	76.6	76.9
$e + \mu$	84.3	84.2	-	83.9	84.4	-	83.9	83.7	-

Table 3.4: Reconstruction efficiencies for direct lepton decay channels after a loose event selection.



(a) w.r.t. η (b) w.r.t. p_T

Figure 3.7: Electron reconstruction efficiency ε w.r.t. p_T and η for all generator level electrons in the PBH $\rightarrow ee$ sample with $M_P = 1500$ GeV and flavour-conservation.

3.5 Background processes

There are many Standard Model (SM) processes which can lead to a final state with two or more leptons. To account for these, fully simulated official datasets will be used. In Table 3.5 all datasets considered in the following feasibility study are listed, whereas a more complete listing with cross sections and luminosities, is given in appendix C. Drell-Yan (DY) is the most dominant background for di-lepton signals, but other important backgrounds are di-jets, W + jets and di-bosons.

3.5.1 Electroweak gauge bosons

Processes including the production and/or exchange of electroweak gauge bosons are the most important background for lepton searches, and depending on the search channel, some may serve as irreducible backgrounds.

Dataset	Comments
Pythia_DY[ee/mumu/tautau]_[X]M[Y]	SM background. DY, $M_{Z/\gamma^*} \in [X, Y]$. Better statistics at high masses.
PythiaZ[ee/mumu/tautau]	SM background. DY, high-statistics sample without mass windows. Better statistics for masses in the vicinity of the Z mass.
J[X]_pythia_jetjet	SM background. Di-jets, excluding top quark, X from 0 to 8.
T1_McAtNlo_Jimmy	SM background. Top pair, next-to-leading-order.
AlpgenJimmyW[e/mu/tau]nuNp[X]	SM background. $W^{\pm}$, X is number of additional jets.
Pythia_MadGraph_Z[ee/mumu/tautau]gamma	SM background. $Z + \gamma$.
Pythia_MadGraph_Wplus[e/mu/tau]nugamma	SM background. $W^+ + \gamma$.
Pythia_MadGraph_Wminus[e/mu/tau]nugamma	SM background. $W^- + \gamma$.
[WW/ZZ/WZ]_Herwig	SM background. Production of di-bosons.
Pythia_photos_diphotons50	SM background. $\gamma\gamma$ with a lower mass bound of 50 GeV.
Pythia_photos_diphotons100	SM background. $\gamma\gamma$ with a lower mass bound of 100 GeV.
PythiaPhotonJet_Unbinned70	SM background. γ + jet with a lower mass bound of 70 GeV.
PythiaPhotonJet_Unbinned280	SM background. γ + jet with a lower mass bound of 280 GeV.
LM1_BM...2Electrons	Signal.
LM1_BM...2Muons	Signal.
LM1_BM...1Z_X	Signal.
LM1_BM...2Comb_b_t_e_mu_tau_W	Signal.
LM1_BM...2Photons	Signal.
LM1_BM...ElectronMuon	Signal.
LM1_BM...Other	Signal.

Table 3.5: List of background and signal datasets used in the feasibility study.

Prompt photon production

The Pythia_photos_diphotons[X] and PythiaPhotonJet_Unbinned[X] samples contain the following processes[4]:

$$\begin{array}{rrcl} 14 & q_i\bar{q}_i & \rightarrow & g\gamma \\ 18 & f_i\bar{f}_i & \rightarrow & \gamma\gamma \\ 29 & q_i g & \rightarrow & q_i\gamma \\ 114 & gg & \rightarrow & \gamma\gamma \end{array}$$

Processes 14 and 29 are the main source of single-photon production in hadron colliders, while process 115,

$$gg \rightarrow g\gamma,$$

is only important in some kinematic regions and is thus not included in the mentioned samples. Photon conversions and photons interacting with the ECAL give an electron background which can be reduced by requiring higher track quality. In Figure 3.8 two Feynman diagrams of photon production are shown.

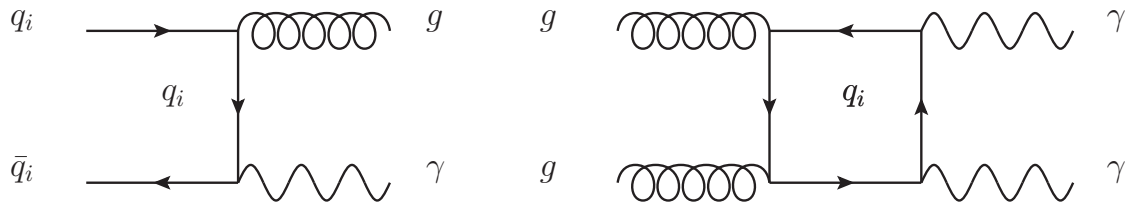


Figure 3.8: Example of single-photon (to the left) and di-photon (to the right) production.

Single $W/Z/\gamma^*$ production

The single massive gauge boson production processes are accounted for in the Pythia_DY[ll], PythiaZ[ll], and AlpgenJimmyW[lν] samples. Drell-Yan processes are described in Pythia by ISUB 1,

$$f_i\bar{f}_i \rightarrow \gamma^*/Z^0(\rightarrow l^+l^-)^5,$$

a $2 \rightarrow 1$ single resonance production. Initial-state radiation is turned on, allowing for additional jets. W production is modeled in ALPGEN [57] by the subprocesses (jproc) 1-25, with the simplest being jproc 1,

$$q_i\bar{q}_j \rightarrow W.$$

This set of processes includes up to four final-state quarks. Gluons are added to the event if the requested jet multiplicity (N $\leq$ 6) exceeds the number of final-state quarks. Explicitly, this means that jproc 1, in the case of two jets corresponds to a final state with two additional gluons, while a process with e.g. three final-state quarks is not allowed. The decay of massive gauge bosons to taus, which again decay to muons or electrons with a probability of about 35%, is a background to be considered as well. In Figure 3.9 two examples of single $W/Z/\gamma^*$ production is shown.

The Pythia cross sections are leading-order (LO), meaning that only the simplest Feynman diagram possible for the given process is considered. In [58] the next-to-next-to-leading-order

[4]The number given before each process corresponds to the respective ISUB in stand-alone Pythia or "pysubs msub" within the Athena interface to Pythia.

[5]The allowed boson decay channels can be controlled by the "pydat3 mdme" switches within the Athena interface to Pythia.

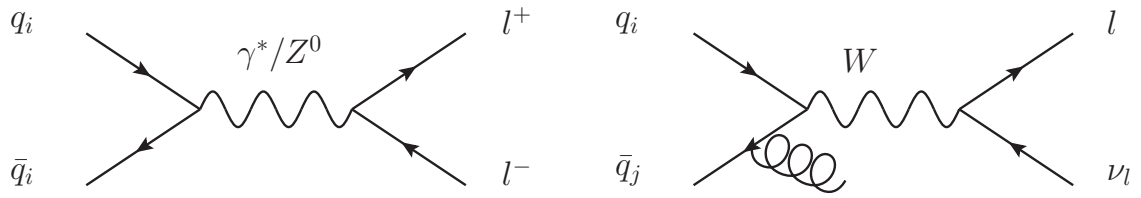


Figure 3.9: The production and decay of a virtual photon or Z to two leptons (to the left) and production and decay of W + jet to lepton + neutrino (to the right).

(NNLO) cross section for single Z/γ^* production was calculated, and the so-called *K-factors* for the binned Drell-Yan samples were extracted, see Table 3.6. A K-factor is the ratio of a higher order estimation and a lower order estimation of the cross section, i.e. to obtain the NNLO cross section, the LO cross-section can be multiplied by the NNLO/LO K-factor. In the following feasibility study, the K-factor of the 1750-2000 GeV binned sample will be applied to the $>$ 2000 GeV binned sample, whereas a K-factor of 1.15 will be applied to the unbinned Drell-Yan samples.

Mass bin (GeV)	K-factor
75-120	1.155
120-250	1.149
250-400	1.113
400-600	1.086
600-800	1.056
800-1000	1.027
1000-1250	0.997
1250-1500	0.951
1500-1750	0.891
1750-2000	0.824

Table 3.6: K-factors to be applied to the binned Drell-Yan samples.

W/Z pair production

The herwig_[di-boson] samples account for ZZ, WZ and WW production. Such a pair can produce two or more leptons in a multitude of ways. Examples of di-boson production are shown in Figure 3.10.

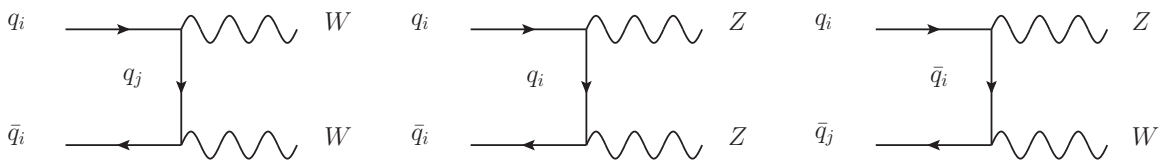


Figure 3.10: Production of WW (to the left), ZZ (in the middle) and WZ (to the right).

W/Z + photon production

These processes are similar to single W/Z production, but an additional photon is produced, and they can be considered as an electroweak correction to the W/Z production. The samples used are produced with MadGraph [59] interfaced to Pythia. In Figure 3.11 two examples of relevant Feynman diagrams are shown.

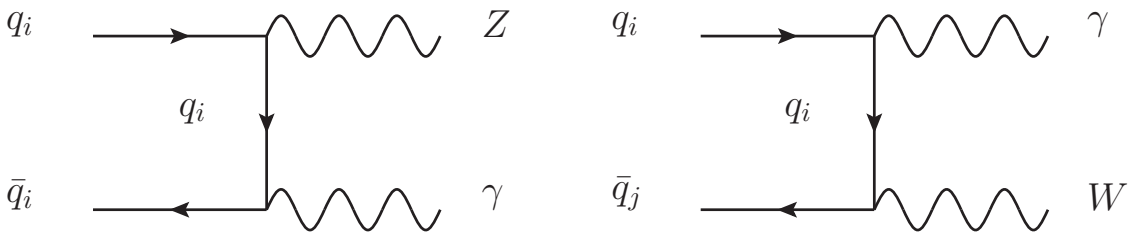


Figure 3.11: Production of $Z + \gamma$ (to the left) and $W + \gamma$.

3.5.2 QCD

A jet is often reconstructed as both a jet and an electron as it may, just as the electron, leave a track in the inner detector and deposit energy in the electromagnetic calorimeter. Imposing hadronic leakage, shower shape and track quality cuts removes most of these jets, but due to the large cross section, QCD processes (see Figure 3.12) are still an important electron background. The copiously produced charged pion, which nearly always decays[6] to a muon and a neutrino, is a considerable low-p_T muon background. Real, isolated electrons and muons may be produced in heavy flavour-decays, whereas punch-through of hadrons into the muon system serves as a fake muon background.

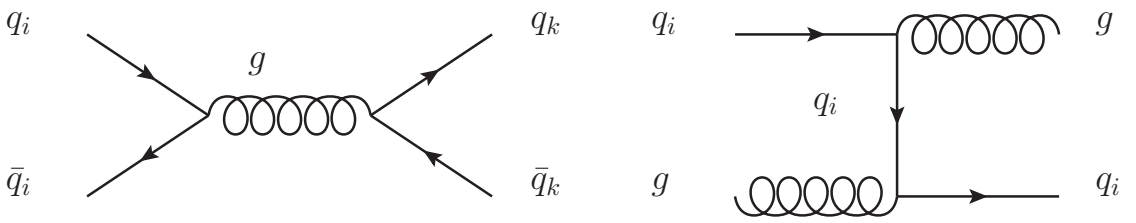


Figure 3.12: Two examples of QCD processes, production of a quark-antiquark pair (to the left) and gluon-quark pair (to the right).

The J[X]_pythia samples contain the following leading-order hard processes with a cross section $\propto \alpha_S^2$:

$$
\begin{array}{lrcl}
11 & q_i q_j & \rightarrow & q_i q_j \\
12 & q_i \bar{q}_i & \rightarrow & q_k \bar{q}_k \\
13 & q_i \bar{q}_i & \rightarrow & gg \\
28 & q_i g & \rightarrow & q_i g \\
53 & gg & \rightarrow & q_k \bar{q}_k \\
68 & gg & \rightarrow & gg
\end{array}
$$

[6]It should be noted that most charged pions charge-exchange in the electromagnetic calorimeter or live long enough to interact with the hadronic calorimeter rather than decaying.

These processes do not give a proper account of masses, and so top-pair production is switched off. There are no higher-order loop corrections, but initial- and final-state QCD radiation is enabled to allow for multi-jet events. In order to account for top-pair production, the T1_McAtNlo sample is used. The MC@NLO [60][61] event generator, based on HERWIG [62], gives a more realistic event generation and computes cross sections to next-to-leading order accuracy. Top quarks decay, in nearly all cases, to a W and a bottom quark, and these may further decay leptonically.

Chapter 4

Feasability study

In general, there are two ways of searching for new physics. One is to do precise measurements and look for deviations from the Standard Model. The other is to test hypotheses for new models and search for events with distinct signatures. In this thesis, the latter will be done. In order to do so, it is essential to know the signatures of the new model and whether the experiment is sensitive. Using fully simulated Release 15 background samples, as well as fast simulated signal samples (see Section 3.3), the signatures of PBH production will be explored in this chapter. A complete list of datasets with cross sections and integrated luminosities is given in Appendix C.1.

The strategy is to justify the event selection on the basis of Monte Carlo before looking at data. Data will not be analyzed until Chapter 5, but if large discrepancies between data and MC are found, it might be necessary to alter the event selection. For the task at hand, the BlackMax sample with $M_P = 1.5$ TeV and flavour-conservation (c2) is used. A Planck mass of $M_P = 1.5$ yields a clearly visible signal assuming an integrated luminosity of 100 pb^{-1}, and it is beyond the limit from Tevatron (see Section 1.2.2). Flavour conserving PBHs can not directly decay to electron + muon and are in that sense more "similar" to Standard Model physics than flavour-violating PBHs, and thus $M_P = 1.5$ TeV and flavour conservation appear as a reasonable starting point. Note that the mixed decay channel $2 \times \{e, \mu, \tau, W, b, t\}$ is in this case reduced to $2 \times \{\tau, W, b, t\}$. More details are given in Table 4.1.

At higher Planck masses, signal and background will become more separated, but the signal will suffer from lower cross sections. The set of conservation laws also plays an important role, and this will be further examined.

	Decay channels of $2 \times \{e, \mu, \tau, W, b, t\}$
c1	$e\tau$, $\mu\tau$, $\tau\bar{\tau}$, bb, tt, bt, W^+W^-, Wt, Wb
c2	$\tau\bar{\tau}$, bb, tt, bt, W^+W^-, Wt, Wb
c3	$e\tau$, $\mu\tau$, $\tau\bar{\tau}$, bb, tt, bt, W^+W^-, Wt, Wb

Table 4.1: The decay channels of the mixed channel for different sets of conservation laws. Remember that c1 is conservation of L and B, c2 includes flavour conservation, while c3 is only conservation of B - L.

4.1 Event Selection

The TeV scale gravity group at ATLAS is mainly focusing on a general multi-object search at high invariant masses [63], i.e. not the two-particle final state scenario, and for that reason their event selection can not be adopted. Some theories proposed to address the shortcomings of the SM contain additional gauge symmetries, and thus additional gauge bosons. These are generically called W' and Z', and they share the properties of the W and Z of the SM, but have a larger mass. At ATLAS, such new gauge bosons are searched for in the $l\nu$ and l^+l^- channels, respectively. They have a large mass and decay to very high-p_T leptons, i.e. the Z' signal definitions can be used as a starting point for PBH analysis. The signal definitions are given in Tables 4.2 and 4.3. These are intended for real data analysis, and since only simulations are considered in this chapter, the final di-muon selection turns out to be somewhat different, see Sections 4.1.2, 4.1.6 and 4.1.7. The event selection for di-electrons remains unchanged however. Note that muons identified with the STACO algorithm will be used in this analysis. The STACO algorithm combines ID and MS tracks using statistical methods, see Section 3.4.2.

Cut type	D3PD variable	see Section
At least one vertex with more than two tracks	vx_n, vx_nTracks	
Trigger	L1_EM14	
Cluster-based electrons (author 1 or 3)	el_author	3.4.1
$\|\eta\|$ less than 2.47 and excluding crack region	el_cl_eta	4.1.2
p_T larger than 25 GeV	el_cl_pt	4.1.2
OQ map	checkOQCluster	4.1.8
Both robust medium	isRobustMedium	3.4.1
Invariant mass > 75	el_tracktheta, el_trackphi, el_cl_E	4.1.2

Table 4.2: The event selection for di-electrons (standard Z' selection for Release 15 reconstructed data and MC, https://twiki.cern.ch/twiki/bin/view/AtlasProtected/CutflowsZprime).

Some of the cuts will be placed right away, see Table 4.4, while the others will be discussed before being possibly applied, see Table 4.5. Triggers with a p_T threshold of $\sim$ 14 GeV in the ECAL and $\sim$ 10 GeV in the MS, respectively, will be used in this chapter. All events are required to fire at least one of these triggers. The p_T, η, vertex, RobustMedium (for electrons), reconstruction algorithm (for electrons) and combined (for muons) cuts will always be placed on reconstructed objects. Electrons are required to be reconstructed with one of the cluster-based algorithms, "author" 1 or 3 in the AOD and D3PD, where 3 is track-based as well. Muon candidates are required to be combined, i.e. have an associated ID track as well as MS track. This rejects muons from e.g. meson decays in the calorimeter. Since one expects a high level of activity in pp collisions, vertex cuts reject events not produced in a pp collision, e.g. beam-gas or beam-beampipe collisions.

4.1.1 A note on resolution

The muon momenta can only be measured by means of tracking, with an accuracy that depends on the bending of the particle. The energy resolution thus increases w.r.t. p_T. This is illustrated in Figure 4.1(a), in which the mean resolution for various p_T ranges have been calculated using all combined generator level muons from the Pythia_DYmumu samples. Electrons, being measured in the calorimeter, do not suffer from this effect, and the resolution actually decreases with

Cut type	D3PD variable	see Section
Primary vertex with more than two tracks	vx_n,vx_nTracks, vx_type, vx_z	4.1.5
Trigger	L1_MU10	
Combined muon	mu_isCombinedMuon	
$\|\eta\|$ less than 1.05	mu_eta	4.1.2
Opposite charge	mu_qoverp	4.1.4
MS and ID hits		4.1.7
p_T larger than 25 GeV	mu_me_qoverp, mu_me_theta	4.1.2
Impact parameters	mu_d0_exPV, mu_z0_exPV	4.1.5
Isolation	mu_ptcone30, mu_me_qoverp, mu_me_theta	4.1.6
Invariant mass > 60	mu_eta, mu_phi, mu_me_qoverp, mu_me_theta	4.1.2

Table 4.3: The event selection for di-muons (standard Z' for Release 15 reconstructed data and MC, https://twiki.cern.ch/twiki/bin/view/AtlasProtected/CutflowsZprimeMuons). The full variable names are obtained by replacing "mu" with "mu_staco".

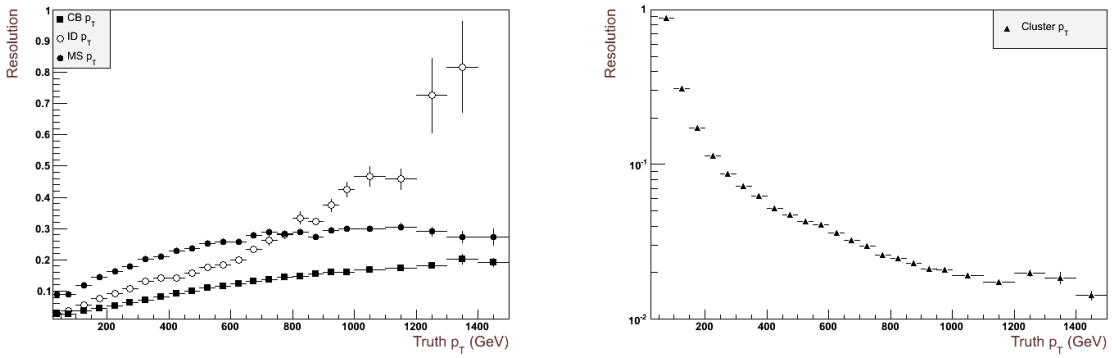
Leptons	cuts
Both	Vertex, trigger, η, p_T, invariant mass
Electrons	Author, RobustMedium
Muons	Combined

Table 4.4: The base event selection.

Leptons	cuts
Electrons	OQ map (see Section 4.1.8)
Muons	Impact parameters, opposite charge, number of hits

Table 4.5: Cuts that will be examined before being possibly applied.

energy, see Figure 4.1(b). The resolution in the muon spectrometer is optimal at $\sim$ 100 GeV, but dominated by fluctuations in the energy loss in the calorimeters at lower values. For large p_T the intrinsic MDT tube accuracy dominates. For this reason, the invariant mass requirement is more generous for muons than for electrons.



(a) Muons (b) Electrons

Figure 4.1: Mean relative resolution of muon tracks (a) and ECAL energy (b). The means are based on all truth (generator level) muons and electrons in the Pythia_DYee/mumu samples (after skimming).

4.1.2 Energy and direction

The Z' cuts are aimed at being close to the cuts used by the WZ working group [64]. For electrons it was decided to use calorimeter information for all purposes except for calculating masses, where one uses calorimeter energy, but track directions θ and ϕ. According to [35], the ID momentum measurement of high p_T electrons is not expected to improve the accuracy of the calorimeter energy measurement.

Due to poor ID alignment in data reconstructed with Release 15, the p_T of the MS track extrapolated to the primary vertex is used instead of the combined p_T. The MS suffers from bad alignment as well, and thus only barrel muons are used, that is, those satisfying $|\eta| < 1.05$. However, these problems are not present in simulation and are fixed in Release 16. Since only simulations are considered in this chapter, the combined muon variables will be used.

A cut on the transverse momentum rejects low energy QCD background. Both electrons and muons are required to have a p_T above 25 GeV, retaining Z boson and higher mass events. For electrons the $|\eta| < 2.47$ cut includes barrel and end-cap electrons, whereas the forward calorimetry is currently not in use. The crack region, defined as $1.47 > |\eta| > 1.52$, is also excluded. Since only simulations are considered in this chapter, a cut on $|\eta| < 2.4$ will be used for muons unless stated otherwise. This is the coverage of the muon triggering system. The energy and direction cuts to be applied are then:

Electrons

- Cluster $p_T > 25$ GeV
- Cluster $|\eta| < 2.47$ and not $1.37 < |\eta| < 1.52$
- $M_{ee} > 75$ GeV, using track directions and cluster energy

Muons

- Combined $p_T > 25$ GeV
- Combined $|\eta| < 2.4$
- $M_{\mu\mu} > 60$ GeV, using combined directions and momentum

4.1.3 Pre-selection

A pre-selection (skimming) of events was applied to all samples. Skimming removes most uninteresting events from the sample files, leaving only a fraction of the total. This is for practical reasons, as the analysis can be quite time-consuming if it is to run over not only interesting events, but also those which anyway will be rejected in the analysis. This is especially effective for large datasets such as pythia_jetet where most events contain very few or no good leptons. The pre-selection is as follows:

Electrons

- Cluster $p_T > 25$ GeV
- Cluster $|\eta| < 2.47$
- Author (reconstruction algorithm) 1 or 3
- RobustLoose

Muons

- Combined $p_T > 25$ GeV
- CombinedMuon
- Combined $|\eta| < 2.4$

Event

- If number of remaining leptons ≥ 2, accept event

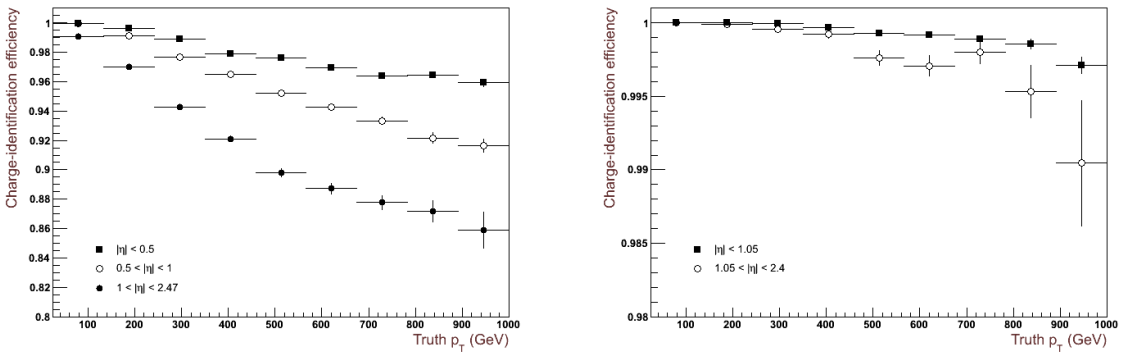
4.1.4 Charge identification efficiency

Due to possible charge-misidentification, opposite charge is not required of the two electron candidates. The charge of an electron can only be measured in the inner detector by means of the bending of the track. The trajectory of a high-p_T electron is less bent by the magnetic field, and the pattern recognition is obscured by electromagnetic showers due to bremsstrahlung. Therefore, the charge-identification efficiency decreases with p_T. The charge-identification efficiency is also highest at low values of $|\eta|$. To illustrate this, the charge measured from the reconstructed track of all RobustMedium electrons from the Pythia_DYee samples (after skimming) was compared to the true charge. In Figure 4.2(a) the charge-identification efficiency of has been plotted w.r.t truth p_T for three ranges of $|\eta|$. In total, 96.0% of all truth $|\eta| < 2.47$ and RobustMedium electrons had the correct charge assigned.

Opposite charge is required when measuring charge asymmetries arising from e.g. a possible heavy gauge boson, but it will at the same time reduce the number of signal events when dominated by high-p_T electrons. Since the charge of fake electrons is assumed to be uncorrelated, requiring opposite charge increases the jet and photon rejection. However, as will soon become evident, the available MC contain very little high-mass fake di-electrons. For this reason, the electron pair candidates are herein not required to have opposite charge. This has to be reconsidered when analyzing real data though.

Since muons are measured in both the inner detector and muon spectrometer, high-p_T muons are less prone to having the wrong sign assigned. This is clearly seen in Figure 4.2(b), where the

efficiency has been plotted w.r.t. truth p_T of all combined muons from the Pythia_DYmumu samples (after skimming). The total charge identification efficiency is 99.9% for truth $|\eta| < 2.4$ and combined muons. Requiring opposite charge thus has very little effect on the signal, and will from now on be required of all di-muon candidates. This rejects 47% and 31% of the di-jet and W + jets background, respectively.



(a) Electrons (b) Muons

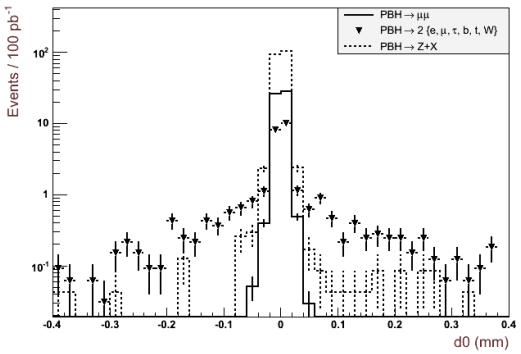
Figure 4.2: Charge identification efficiency for electrons and muons in the Pythia_DYee/mumu samples (after skimming).

4.1.5 Impact parameters and track match

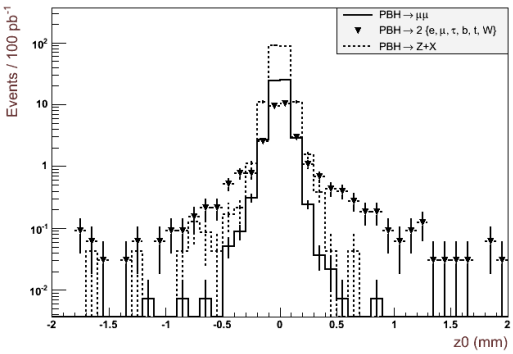
Due to cosmic muons constantly traversing the detector material, there is need of more vertex-specific cuts for muons than for electrons. To suppress events where the detector triggered on cosmic muons, there must be at least one primary vertex (vertex type 1 in the AOD and D3PD) with at least two tracks. The muon track origin is required to be in close vicinity of the primary vertex, that is, at an absolute z position of no more than 200 mm. To further suppress cosmic muons, longitudinal (z_0) and transverse (d_0) impact parameter w.r.t. to primary vertex is required to be less than 0.2 mm and 1 mm, respectively. The impact parameter cuts also reject 33% of the remaining di-jet background.

In Figure 4.3 the impact parameters for PBH muons have been plotted. While the prompt di-muon and $Z + X$ samples give clean peaks, the mixed decay channel shows more of a tail. This is possibly due to muons originating from the relatively long-lived b-quark. It can further be seen that the d_0 cut removes most of the tail in the z_0 distribution, but not the other way around. If the tails are due to b-quark decay, the sphericity of BH decays in general implies that the mean value of z_0 and d_0 for muons associated with PBH b-quarks is fairly equal. Since the d_0 cut is stricter than the z_0 cut, it thus comes as no surprise that the d_0 cut removes most of the tail in the z_0 distribution.

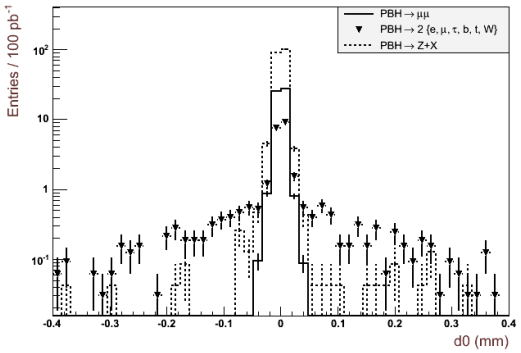
The χ^2 is a measure of how well the inner detector and muon spectrometer tracks match, and a low value indicates a good matching. In Figure 4.4 the χ^2 distributions for PBH, Drell-Yan and di-jet muons have been plotted. Although there are no requirements on track matching in the event selection, it is reassuring that muons from PBHs have relatively good matching between their ID and MS track. Judging from the plots, the χ^2 distribution for signal and background distributions is very similar, which is just as expected. Henceforth, no cut will be applied on the track matching, but the impact parameter cuts are applied unless stated otherwise.



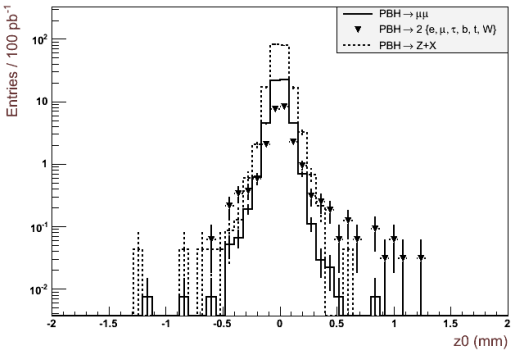
(a) Signal $d0$



(b) Signal $z0$

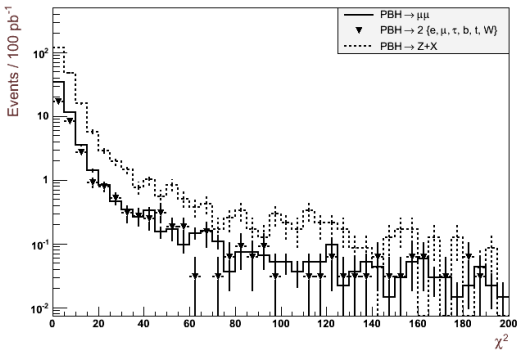


(c) Signal $d0$ after $z0$ cut

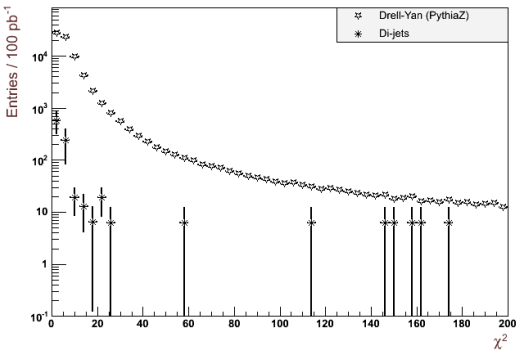


(d) Signal $z0$ after $d0$ cut

Figure 4.3: Impact parameter distributions of the two muon candidates.



(a) Signal χ^2



(b) Background χ^2

Figure 4.4: The χ^2 distribution of the two muon candidates.

4.1.6 Isolation

Isolation cuts are meant to reduce the contamination of jets reconstructed as electrons and muons originating from e.g. semi-leptonic decays of heavy flavour hadrons. If the amount of activity in the vicinity of the lepton exceeds some fixed limit, the lepton is rejected. Whereas there is no isolation cut in the event selection for electrons, the ptcone30 variable is used for muons. This is defined as the p_T-sum of all other tracks within a $\Delta R = 0.3$ cone, or in mathematical form,

$$p_T^{\Delta R<0.3} = \sum_{\Delta R<0.3} p_T.$$

In order to allow more activity in the vicinity of highly energetic particles, due to effects such as bremsstrahlung, the ptcone30 variable is normalized to the muon p_T.

After adding the impact parameter cuts, one can see from Figure 4.5(a) that a significant amount of the muon-pair candidates from PBH Zs are within each others $\Delta R = 0.3$ cone. In that case, it is clear that the ptcone30 normalized to the muon p_T, see Figure 4.5(b), has a value much greater than the upper limit of 0.05.

It might be an idea to use the etcone30 variable instead. This is a measure of the transverse energy measured in the calorimeters within a $\Delta R = 0.3$ cone. Due to deflection of charged particles in the ECAL material, electrons rapidly lose energy to bremsstrahlung radiation. The total radiated power goes as m^{-4} or m^{-6} in the two limiting cases $a \perp v$ or $a \parallel v$, where a is the acceleration and v the speed of the particle. Because of the strong mass dependence, muons lose little energy in the ECAL. As one can see in Figure 4.5, the etcone30 normalized to the muon E_T greatly enhances the isolation, but still $\sim 15\%$ of the signal is lost if a cut at 0.05 is placed.

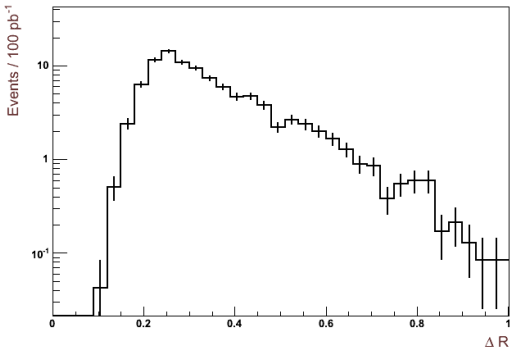
Placing this cut fully suppresses what is left of the MC di-jet background after requiring opposite charge and impact parameters. However, as will soon become evident, the available di-jet MC contain no high-mass or highly boosted di-muons. Thus, to keep the signal acceptance as high as possible, no isolation cuts will be placed in this chapter. There are several other isolation variables with narrower or wider cones, and the need for an isolation cut must be studied later when analyzing real data.

4.1.7 ID and MS hits

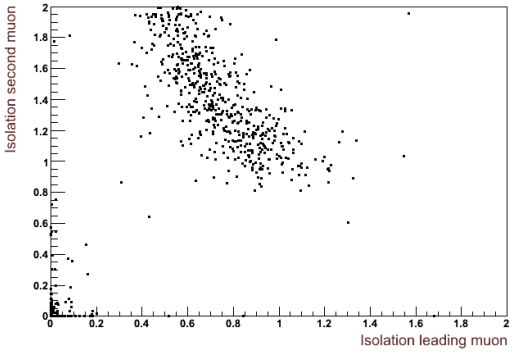
To ensure a satisfactory reconstruction of invariant mass etc., one must apply some requirements on the number of hits in both the inner detector and muon spectrometer. These are listed in Table 4.6. Precise measurements in the bending plane are performed by the MDTs, and thus at least three hits in each barrel (B) layer is desired; inner (BI), middle (BM) and outer (BO). To get a satisfactory measurement of the ϕ angle, at least one phi-hit in at least two RPC layers is required. In Release 16, where the full η coverage of the muon trigger is utilized, these cuts are extended to the end-cap (E) layers as well.

The BEE (Barrel End-cap Extra) and BIS78 chambers are special chambers shaped to fit between magnet coils, support structures etc. The BEE chamber is located at the outer circumference of the end-cap cryostat, where it measures muons passing from the barrel to the end-cap [35]. Due to bad alignment, tracks with hits in the BEE and BIS78 chambers are vetoed.

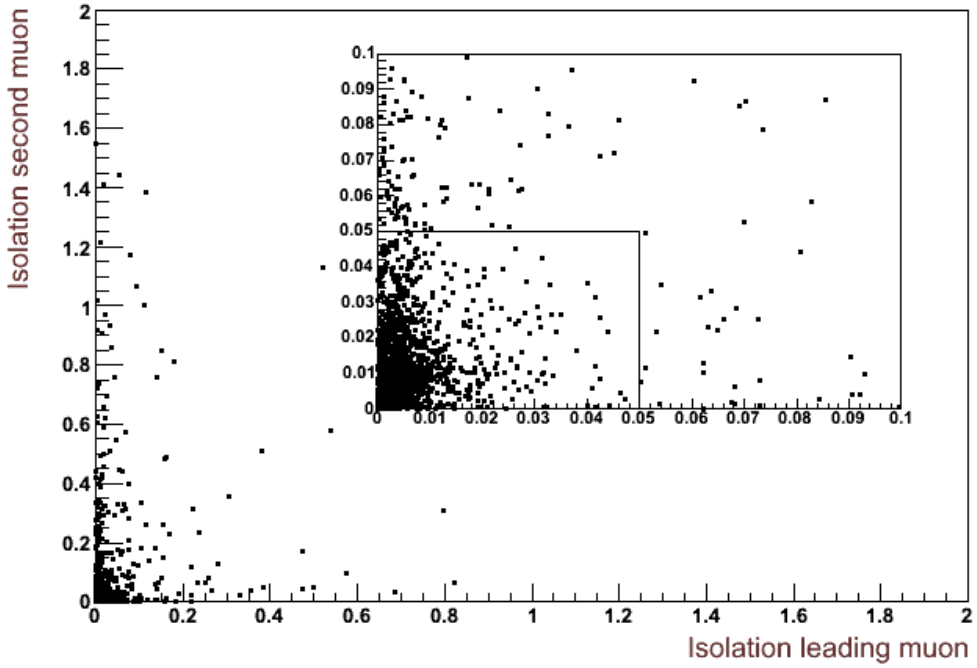
These requirements are specific for barrel muons, that is, muons satisfying $|\eta| < 1.05$. Hence, only barrel events are considered in this section. By imposing the hits requirements as well, the signal in the barrel is reduced by 25%, 21% and 16% for the prompt, boosted Z and mixed decay channel di-muons, respectively. This is clearly a large, but necessary, signal reduction. In combination with the increasing resolution of the muon spectrometer, accepting badly measured muons increases the risk of discovering a fake signal. However, only simulations are studied in this chapter, and it is of interest to utilize the full η-coverage of the MS. Thus, the hits



(a) ΔR distance for muon-pairs



(b) ptcone30/p_T



(c) etcone30/p_T

Figure 4.5: ΔR (a) and isolation distributions for reconstructed muon-pairs in the $Z + X$ sample. The inset in (c) is a zoom of the lower-left corner if the main plot.

Cut type	D3PD variable	requirement
Barrel MDT stations	nMDTBIHits, nMDTBMHits, nMDTBOHits	At least three hits in each MDT layer.
Veto BIS78 and BEE	nMDTBIS78Hits, nMDTBEEHits	No hits
RPC layers	nRPCLayer1PhiHits, nRPCLayer2PhiHits, nRPCLayer3PhiHits	At least one ϕ hit in at least two RPC layers.
Pixel	nPixHits	At least one hit.
SCT	nSCTHits	At least four hits.
Pixel + SCT	nPixHits, nSCTHits	At least six hits in total.

Table 4.6: Muon hits requirements.

requirements will be omitted for now, but the study of these will be resumed in the next chapter when analyzing real data.

The number of ID and MS hits in the different layers have been plotted for the signal only in Figures 4.6, 4.7 and 4.8. Note that the purpose of the hits cuts are not to suppress background, but rather to reject badly measured muons, and thus the distributions in background and signal are more or less the same.

4.1.8 Object Quality maps

The Object Quality maps (OQ maps) includes the fiducial cuts around dead Front-End Boards (FEBs) in the LAr calorimeter and spot the problematic high voltage (HV) regions. The newest map will be used in this chapter, namely Map 166497.

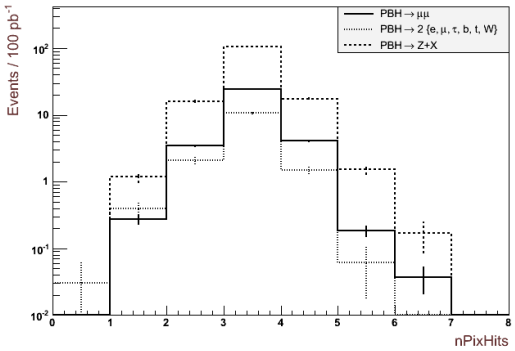
The drift speed of electrons or ions in the LAr gaps depends on the electric field, which typically has a value of 1 kV/mm. For each sub-detector a HV setting is applied, and for the HV sectors with non-optimal behaviour, solutions were implemented in order to recover the corresponding region. The result is that signals from cells seeing reduced high voltage can still be well reconstructed by using a correction scale factor. This is applied online at the energy reconstruction level [65].

The cell signals are read out through 1524 FEBs. They perform analog processing (amplification and shaping), store the signal while waiting for L1 trigger decision and digitize the accepted signals. At the end of September 2009, 1.3% of the cells had problems. The major cause of these cells not being read out is malfunctioning FEBs on which the active part of the optical transmitter to the Readout Driver (ROD) has failed.

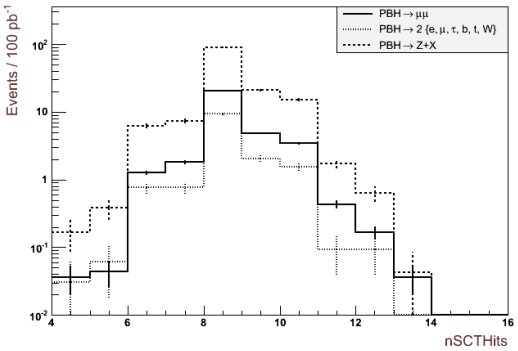
The OQ map utility allows the user to check if there is a problem in the region of the detector covered by the cluster associated to an electron or photon. The function checkOQClusterElectron takes three arguments, run number, η and ϕ, and the possible outputs are:

- 1 - if there are no problems
- 2 - if there is a non-nominal HV setting
- 3 - if there is a major problem (dead FEB, dead HV, ...)

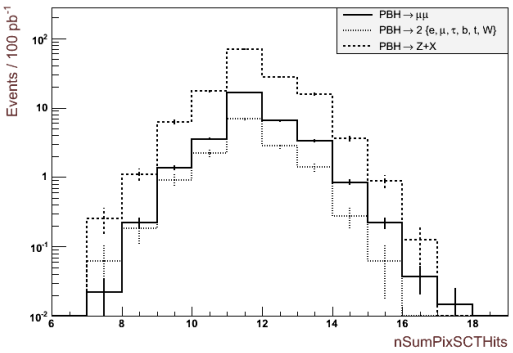
If the output given is 3, the object is rejected. In this analysis, the OQ map check will be applied at the end when the two electron candidates has been chosen. If the check returns 3 for one or both of the electrons, the event is rejected.



(a) Pixel

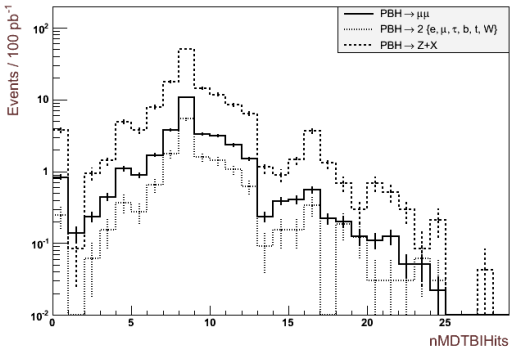


(b) SCT

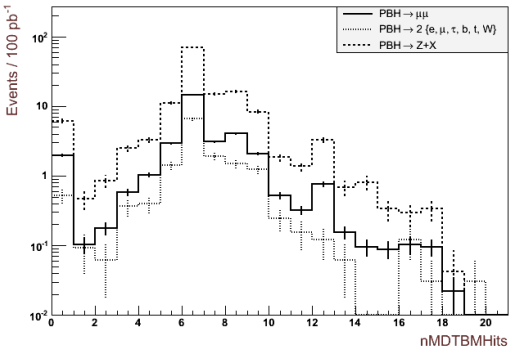


(c) Pixel + SCT

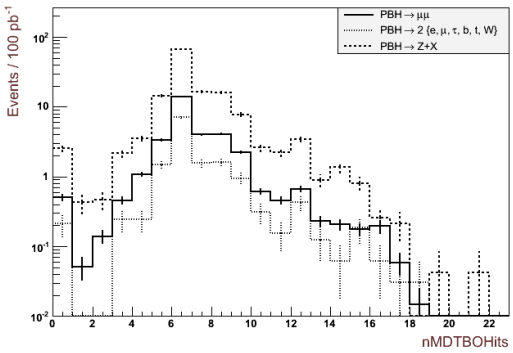
Figure 4.6: Number of hits in Pixel, SCT and Pixel + SCT after selecting only muons with $|\eta| < 1.05$.



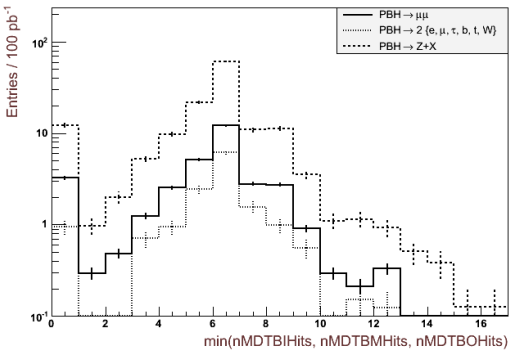
(a) MDT, inner layer



(b) MDT, middle layer

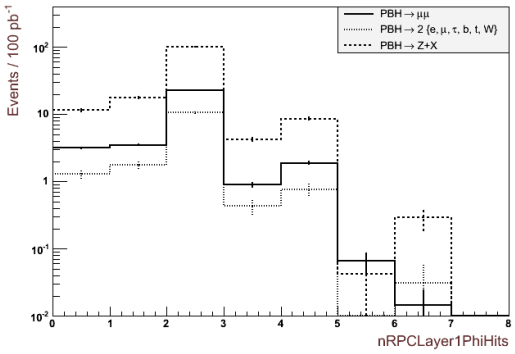


(c) MDT, outer layer

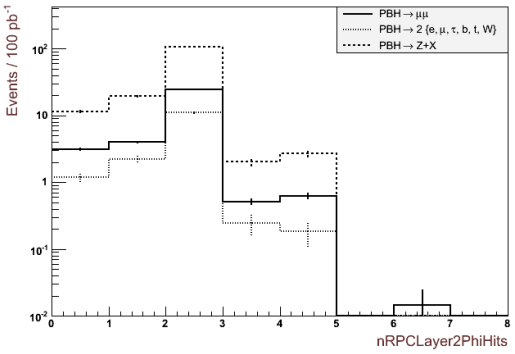


(d) Least number of hits in a MDT barrel layer

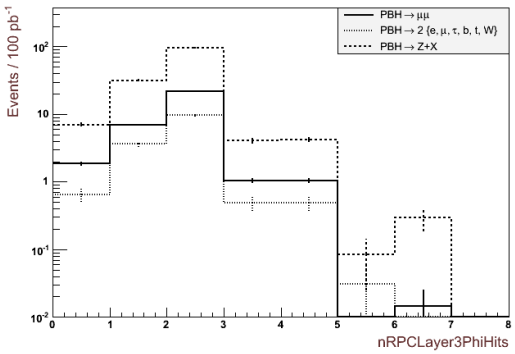
Figure 4.7: Number of MDT hits in each barrel layer and after selecting only muons with $|\eta| < 1.05$.



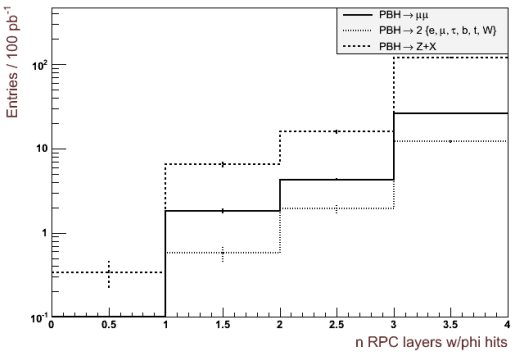
(a) RPC, layer 1



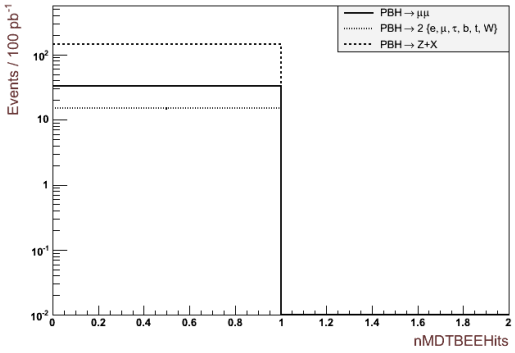
(b) RPC, layer 2



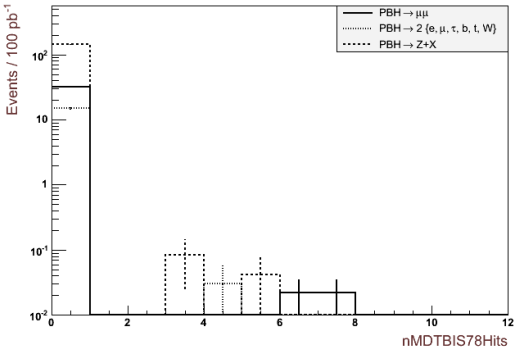
(c) RPC, layer 3



(d) Number of RPC layers with ϕ hits



(e) BEE



(f) BIS78

Figure 4.8: Number of hits in the RPC, BEE and BIS78 chambers and after selecting only muons with $|\eta| < 1.05$.

4.2 Discriminating signal from background

High-mass intermediate states are rare within the Standard Model. For example, the cross section for Z/γ^* production with an invariant mass above 2 TeV is in the order of 10^{-5} pb. Thus, a good discriminator is the invariant mass of the di-leptons. However, if the di-leptons originate from a Z emitted by a PBH, their invariant mass is still the mass of the Z. The pole mass of the Z is 91.1876 GeV, but the momentum of a PBH Z is far above the average Z produced in a SM event.

In the following, various distribution will be shown and discussed. The event selection is as follows: if there are at least two leptons in the event satisfying the cuts in Table 4.7, the event is stored and the two highest-p_T leptons are chosen as the di-lepton.

Leptons	cuts
Both	Vertex, trigger, η, p_T, invariant mass
Electrons	Author, RobustMedium, OQ map
Muons	Impact parameters, combined, opposite charge

Table 4.7: Cuts applied on electrons and muons in the following.

4.2.1 Mass distributions

For both electrons and muons two signal-enriched-region can be seen in Figures 4.9 and 4.10. The first is around the mass of the Z and dominated by Drell-Yan. The other extends from slightly below the PBH production threshold and to a few TeV, rising well above the SM backgrounds. In addition, being a combination of several decay channels, the mixed decay channel has a rather flat mass distribution. The two signal-enriched regions correspond to Z and prompt di-lepton decay. Within the latter, there is a small contribution from prompt di-photon decay as well. Due to low statistics, the simulated di-jet background for both muons and electron has some nasty uncertainty bars.

Only di-lepton events from one of the signal enriched-regions are selected in the following. As the energy resolution in the tracking systems is inferior to that of the ECAL at high energies, it is appropriate to allow larger mass windows for di-muons. The mass windows are defined as follows:

Di-electrons

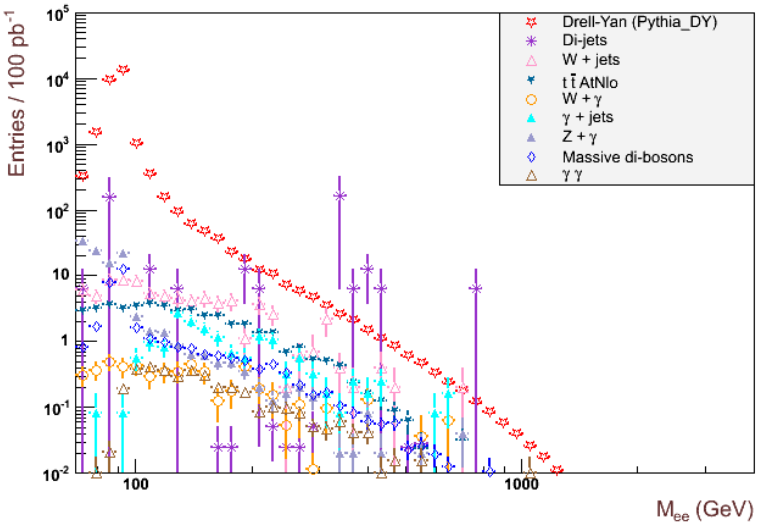
High-mass: > 1500 GeV
Z-mass: $\in (80, 100)$ GeV

Di-muons

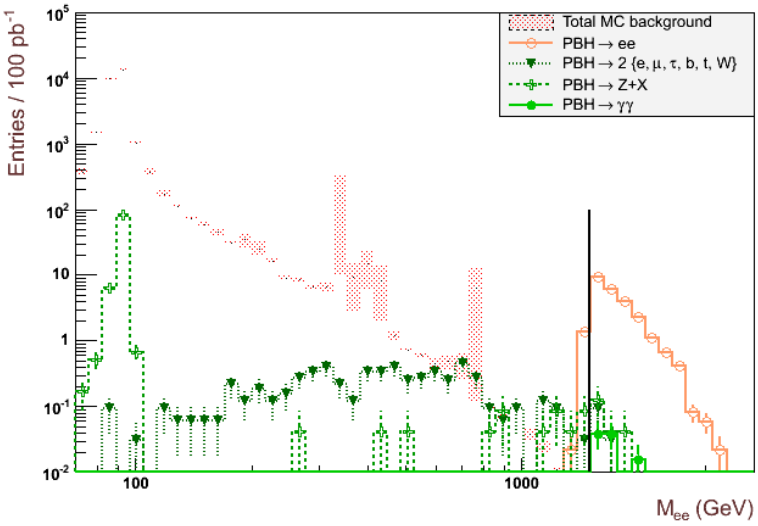
High-mass: > 1200 GeV
Z-mass: $\in (75, 105)$ GeV

These are not optimized in any way, but just chosen by eye. It is of interest to keep the signal acceptance high, but at the same time achieve a good signal-to-background ratio, i.e. in the order of $10^2 - 10^3$ in the high-mass window. In Figure 4.11 it can be seen that the $Z + X$ sample has long tails in the di-muon invariant mass spectrum, and that the chosen Z-mass window for di-muons is in the vicinity of where signal and $t\bar{t}$ background cross. For di-electrons, the $Z + X$ sample has a tail in the lower end of the spectrum and a cut-off at about 100 GeV.

Relative to the base selection, the $PBH \to ee$ and $PBH \to \mu\mu$ acceptance in the high-mass window is 98.1% and 98.6%, respectively. In the Z-mass window, the relative acceptance is

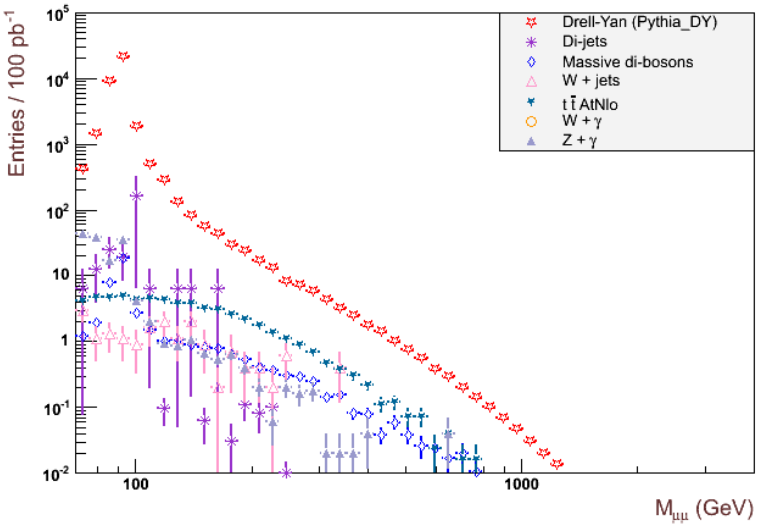


(a) SM background M_{ee}

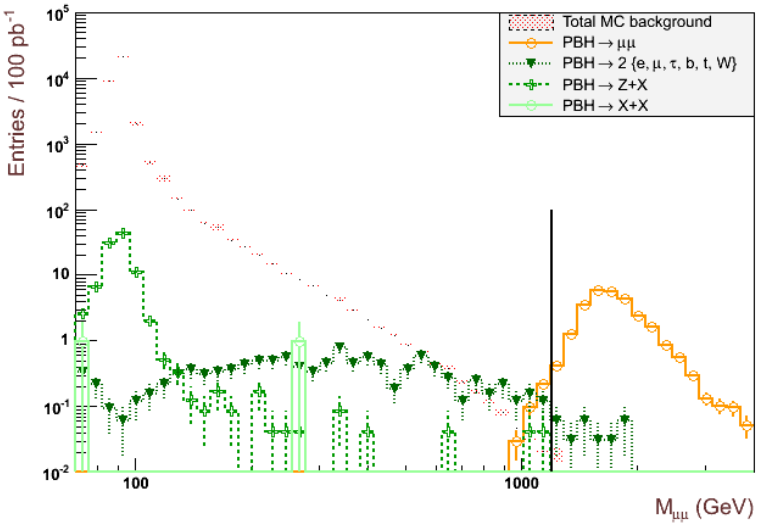


(b) Signal M_{ee} for $M_P = 1.5$ TeV and flavour conservation (c2)

Figure 4.9: Invariant mass distribution of the di-electron candidates, M_{ee}, where the line marks 1.5 TeV. The background in (b) is the sum of the background in (a).



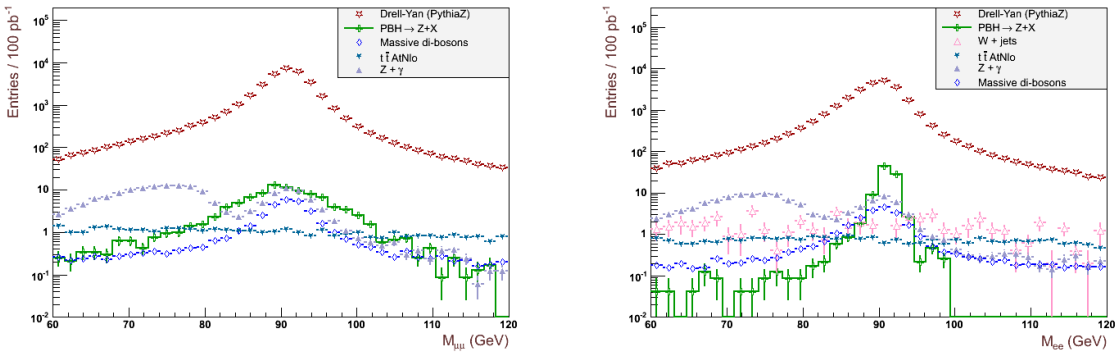
(a) SM background $M_{\mu\mu}$



(b) Signal $M_{\mu\mu}$ for $M_P = 1.5$ TeV and flavour conservation (c2)

Figure 4.10: Invariant mass distribution of the di-muon candidate, $M_{\mu\mu}$, where the line marks 1.2 TeV. The background in (b) is the sum of the background in (a).

98.9% and 91.8% for $PBH \to Z + X \to ee$ and $PBH \to Z + X \to \mu\mu$, respectively. It should be noted that practically all background shown in the plots to come belong to the Z-mass window, while the prompt di-leptons from PBHs purely belong to the high-mass window. The mixed decay channel sample gives small contributions to both.



(a) $M_{\mu\mu}$ (b) M_{ee}

Figure 4.11: Invariant mass distribution of the di-lepton candidates in the Z-mass windows. For the sake of appearance, i.e. due to large uncertainty bars, di-jets, W + jets (for muons) and some less important backgrounds have been left out of the plots. These are small compared to the ones shown.

4.2.2 Angular distributions

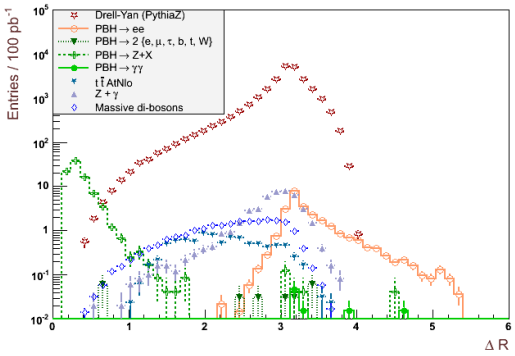
In Figure 4.12 several angular distributions in the di-lepton channels have been plotted. As expected, the distributions are very similar for di-electrons and di-muons. From the η distributions it can be seen that the PBH di-leptons have a more rounded shape than the flatter SM background. This is an indication of the PBH decay products being more transverse. In the ΔR and $\Delta\phi$ distances there is evidence of highly boosted Zs from PBHs. A typical Drell-Yan di-lepton pair is back-to-back, that is, separated by π in R and ϕ. In a PBH Z event, however, half of the total energy is taken by the other particle X such that the Z is boosted in both the z direction and the $x - y$ plane.

4.2.3 Kinematic distributions

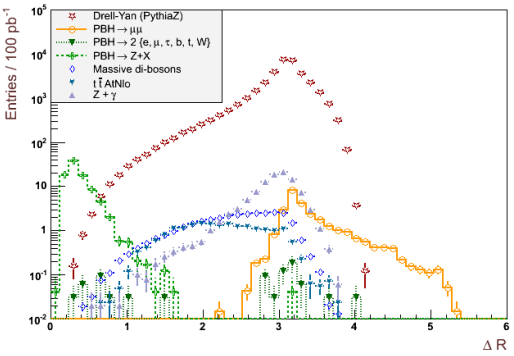
In Figure 4.13 several distributions related to kinematics have been plotted. In the p_T-distributions there are tails extending to several TeV in the di-muon channel. This is absent in the di-electron channel and is due to the inferior momentum resolution of the tracking systems. Otherwise, the di-muon and di-electron distributions are more or less the same. One sees that the p_T distributions clearly separate signal from background, showing that ATLAS is definitively sensitive to high-mass states. The Lorentz factor, which is shown for electrons in Figure 4.13(f), indicates a highly boosted PBH Z, but the di-lepton p_T stands out as being the best discriminator between PBH and SM Z production.

4.2.4 A small note on uncertainty analysis for efficiencies

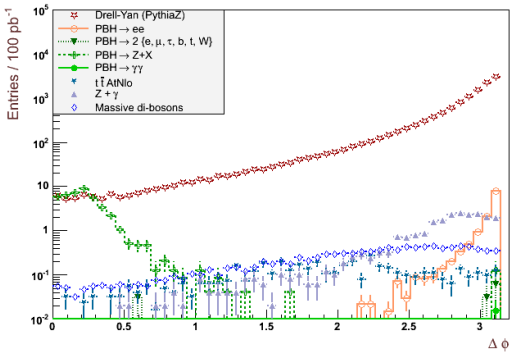
The selection efficiency for a Monte Carlo sample is typically taken by dividing the number of *selected* events m by the number *generated* N. The number of selected events m is usually treated as a binomially distributed variable and N as a constant. The probability of selecting



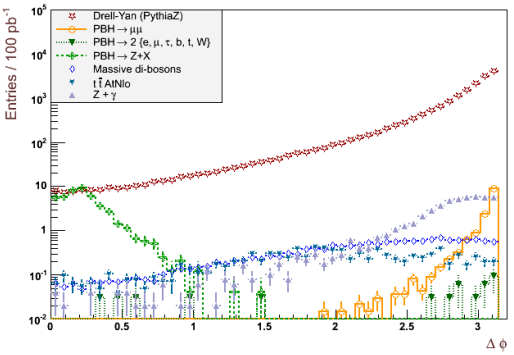
(a) Di-electron ΔR distance



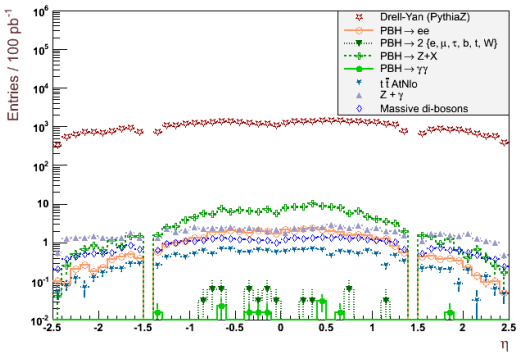
(b) Di-muon ΔR distance



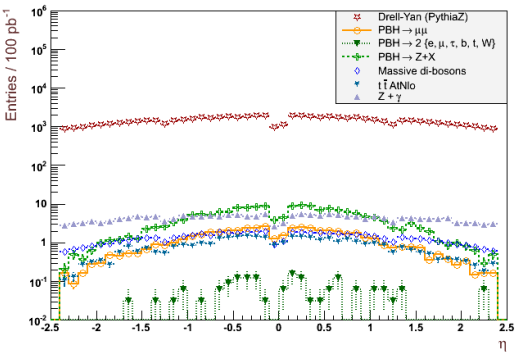
(c) Di-electron $\Delta\phi$ distance



(d) Di-muon $\Delta\phi$ distance

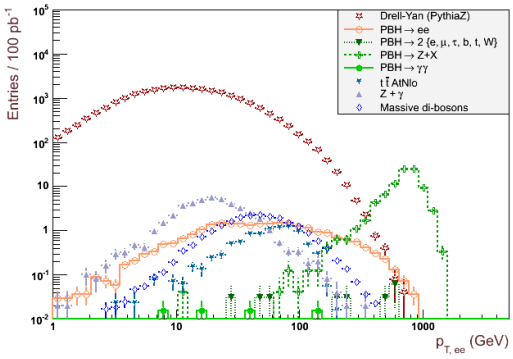


(e) η of the two electron candidates

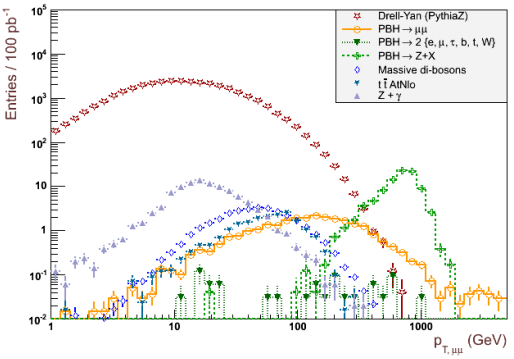


(f) Di-muon η of the two muon candidates

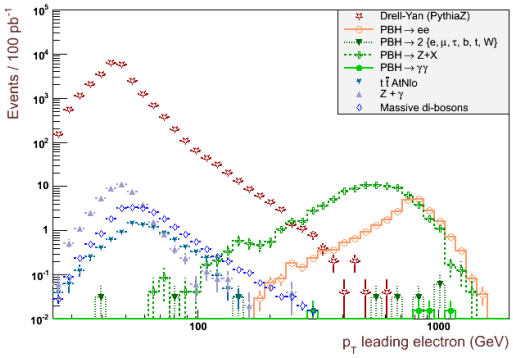
Figure 4.12: Angular distributions in the same-flavour di-lepton channels.



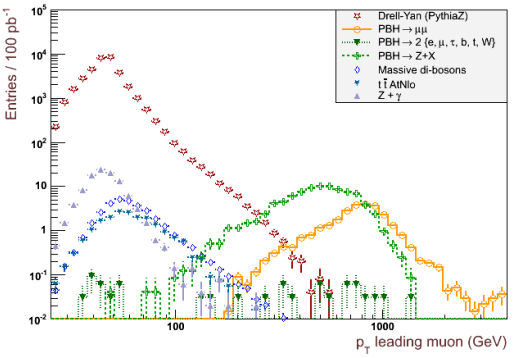
(a) p_T of the di-electron



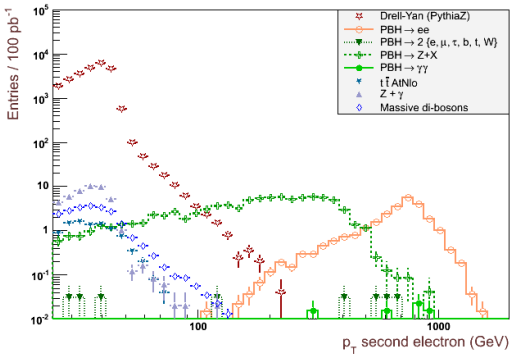
(b) p_T of the di-muon



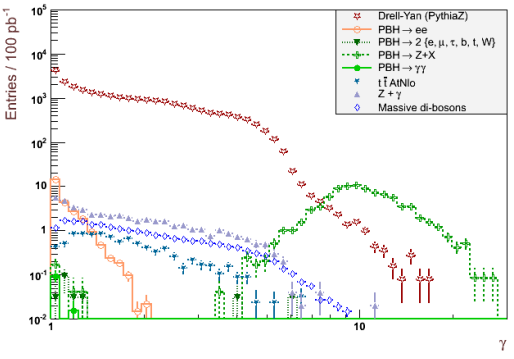
(c) p_T of the leading electron candidate



(d) p_T of the leading muon candidate



(e) p_T of the second electron candidate



(f) Lorentz factor of the di-electron

Figure 4.13: Kinematic distributions in the same-flavour di-lepton channels.

m events out of N is then given by

$$P(m; N, \epsilon) = \frac{N!}{m!(N-m)!}\epsilon^m(1-\epsilon)^{N-m}, \tag{4.1}$$

where ϵ is the unknown efficiency. An estimate of the efficiency is given by

$$\hat{\epsilon} = \frac{m}{N}, \tag{4.2}$$

and given that the variance of the binomially distributed m is

$$V[m] = N\epsilon(1-\epsilon), \tag{4.3}$$

the variance of the estimator $\hat{\epsilon}$ is

$$V[\hat{\epsilon}] = V\left[\frac{m}{N}\right] = \frac{1}{N^2}V[m] = \frac{\epsilon(1-\epsilon)}{N}. \tag{4.4}$$

Since ϵ is unknown, an *estimator* of the variance of the estimator is needed, and this can be taken to be

$$\hat{V}[\hat{\epsilon}] = \frac{\hat{\epsilon}(1-\hat{\epsilon})}{N} = \frac{m(1-m/N)}{N^2}. \tag{4.5}$$

The estimate of the standard deviation is then simply giving by taking the square root,

$$\hat{\sigma}[\hat{\epsilon}] = \frac{\sqrt{m(1-m/N)}}{N}. \tag{4.6}$$

This ensures that

$$\hat{\epsilon} - \hat{\sigma}[\hat{\epsilon}] > 0, \tag{4.7}$$

$$\hat{\epsilon} + \hat{\sigma}[\hat{\epsilon}] < 1, \tag{4.8}$$

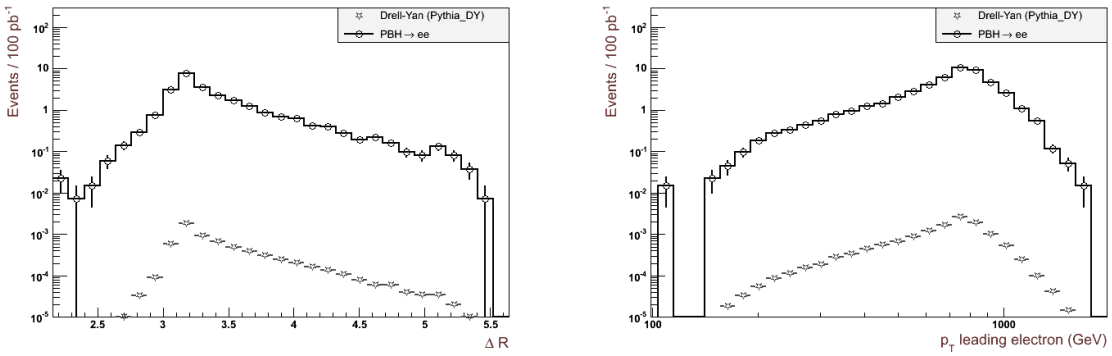
always.

In ROOT, the statistical uncertainties are calculated assuming that the number of accepted events follows a Poisson distribution. This holds in the limit when there is a large number of possible events, each of which is rare. Although the difference is small, m is more correctly treated as a binomially distributed variable. All statistical uncertainties in this chapter are calculated from Eq. 4.6.

4.2.5 High-mass window

Only high-mass di-leptons are selected in this section. Due to the low statistics of the unbinned Drell-Yan samples in this high-mass regime, they are replaced by the binned Drell-Yan samples. As indicated in Figure 4.14, where the ΔR and leading electron p_T have been plotted, Drell-Yan and PBHs are indistinguishable. However, this is not a problem, as the cross section for Drell-Yan is 3-4 orders of magnitude below PBHs.

In the high-mass di-electron channel there are a couple of highly uncertain contributions from di-jet background and $Z + X$ signal. This is reflected in the large standard deviation in the signal-to-background ratio, see Table 4.8. In the high-mass di-muon channel, there is no jet contamination nor $Z + X$ signal, and apart from Drell-Yan, the only background is a small contribution from massive di-boson production, see Table 4.9. The signal-to-background ratio in the high-mass di-electron channel is an impressive 2070±600. As a consequence of the lower mass bound in the high-mass di-muon channel, allowing more Drell-Yan to pass, the signal-to-background ratio is here 680±81.



(a) ΔR (b) Leading electron p_T

Figure 4.14: ΔR and leading electron p_T in the di-electron channel in the high-mass window.

Process	events / 100 pb^{-1}	± stat. uncertainty
Drell-Yan ($e^+e^-/\tau^+\tau^-$)	0.006568	±0.000028
Di-jets	0.0050	±0.0035
γ + jets	0.00084	±0.00048
PBH $\rightarrow ee$	25.23	±0.24
PBH $\rightarrow Z + X$	0.25	±0.10
PBH $\rightarrow 2\{e, \mu, \tau, b, t, W\}$	0.123	±0.061
PBH $\rightarrow \gamma\gamma$	0.116	±0.030

Table 4.8: Number of expected di-electron events with invariant mass above 1.5 TeV. The signal model is $M_P = 1.5$ TeV and flavour conservation (c2).

Process	events / 100 pb^{-1}	± stat. uncertainty
Drell-Yan ($\mu^+\mu^-/\tau^+\tau^-$)	0.03451	±0.00060
Massive di-bosons (W^+W^-/WZ)	0.0060	±0.0048
PBH $\rightarrow \mu\mu$	27.16	±0.23
PBH $\rightarrow 2\{e, \mu, \tau, b, t, W\}$	0.276	±0.092

Table 4.9: Number of expected di-muon events with invariant mass above 1.2 TeV. The signal model is $M_P = 1.5$ TeV and flavour conservation (c2).

4.2.6 Z-mass window

Before any further cuts, the Z-mass window is dominated by background. However, several ways of reducing it can be extracted from Figures 4.12 and 4.13. Many of these variables are correlated, and it is thus preferable to cut on only one of them. The di-lepton p_T is a good candidate as it speaks to both the boost of the Z and the high sphericity of PBH decay. Here, PBH Z and Drell-Yan meet at a low point, beyond which Drell-Yan falls quickly. In Figure 4.15 it can be seen that a cut at 350 GeV is suitable for this particular case, $M_P = 1.5$ TeV and flavour-conservation. The same cut is applied to both di-muons and di-electrons.

In Tables 4.10 and 4.11 the expected number of events for each process is listed. The signal-to-background ratios are 43.1±6.1 and 30.9±3.7 for boosted di-electrons and di-muons, respectively. With 100 pb^{-1} of data one expects 176 Z candidates, of which 172 are PBH Zs. It is thus safe to conclude that this signal definitively has discovery potential.

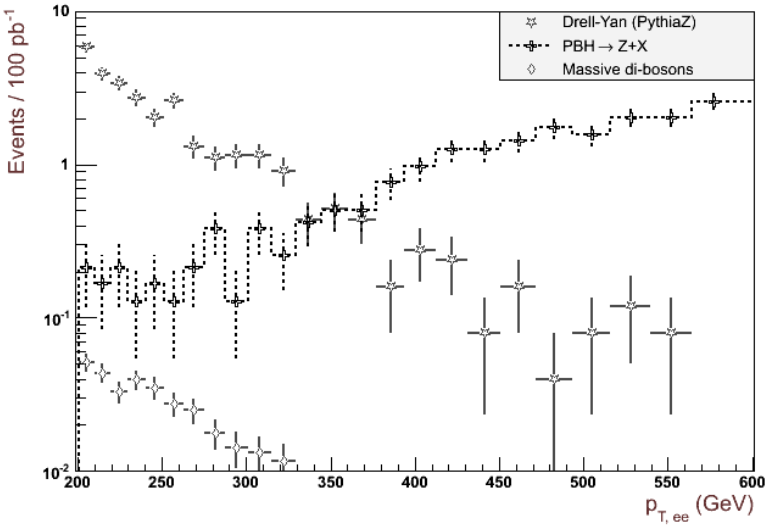


Figure 4.15: Di-electron p_T

Process	events / 100 pb^{-1}	± stat. uncertainty
Drell-Yan (e^+e^-)	1.94	±0.28
Di-jets	0.000012	±0.000012
Massive di-bosons (W^+W^-/WZ)	0.0499	±0.0066
PBH → ee	85.6	±1.7
PBH → $2\{e, \mu, \tau, b, t, W\}$	0.031	±0.031

Table 4.10: Number of expected di-electron events in the Z-mass window after di-lepton p_T cut. The signal model is $M_P = 1.5$ TeV and flavour conservation (c2).

Dataset	events / pb^{-1}	$\pm$ stat. uncertainty
Drell-Yan ($\mu^+\mu^-$)	2.72	$\pm$0.33
$Z(\rightarrow \mu^+\mu^-) + \gamma$	0.020	$\pm$0.020
Massive di-bosons (W^+W^-/WZ)	0.0493	$\pm$0.0067
PBH $\rightarrow \mu\mu$	86.0	$\pm$1.7
PBH $\rightarrow 2\{e, \mu, \tau, b, t, W\}$	0.153	$\pm$0.068

Table 4.11: Number of expected di-muon events in the Z-mass window after di-lepton p_T cut. The signal model is $M_P = 1.5$ TeV and flavour conservation (c2).

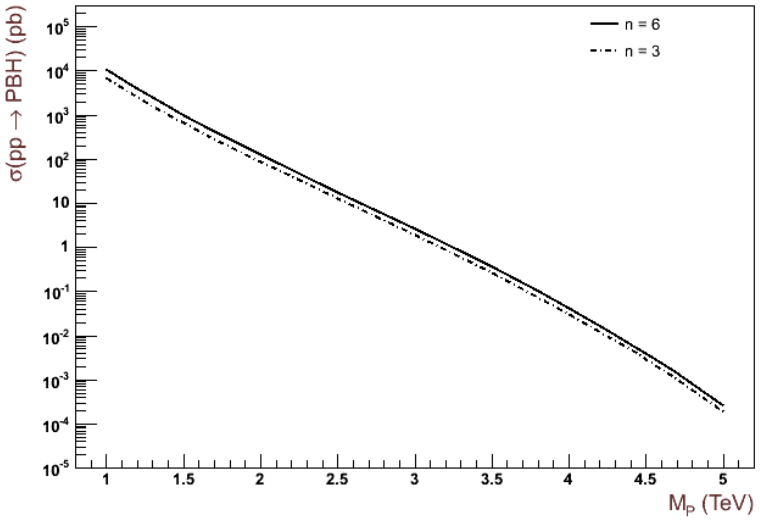


Figure 4.16: PBH production cross section at $\sqrt{s} = 7$ TeV as a function of Planck mass for three and six large extra dimensions n.

4.3 Higher Planck masses

From Eq. 1.38 it can be deduced that the parton-parton cross section for semi-classical BH formation decreases with the Planck mass as

$$\sigma_{ij\to BH}(s) \propto \frac{1}{M_{DL}^{2\left(1+\frac{1}{n+1}\right)}}. \tag{4.9}$$

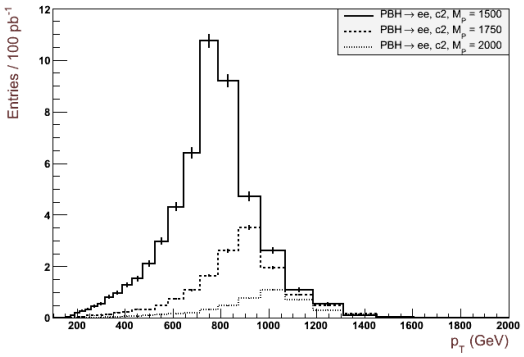
Up to this point, production of flavour-conserving PBHs has been studied at a Planck mass of 1.5 TeV. If the Planck mass is increased to 1.75 TeV, the parton-parton cross section decreases by $\sim 30\%$, and the assumed PBH threshold is shifted to 1.75 TeV. Hence, partons with higher energies are required to produce PBHs, and this further reduces the probability of producing PBHs. Taking the parton PDFs into consideration, the production cross section is reduced by $\sim 65\%$ and $\sim 87\%$ if the Planck mass is increased to 1.75 TeV and 2 TeV, respectively. The PBH production cross section at $\sqrt{s} = 7$ TeV as a function of Planck mass is shown in Figure 4.16.

The invariant mass and p_T distributions for the same-flavour di-lepton channels have been plotted for different Planck masses in Figure 4.17. The plots, being linear on the y-axis, do not give a proper picture of the distribution shapes for higher Planck masses. They seem to flatten as the Planck mass increases, but this is not true. However, the idea is to illustrate how a higher Planck mass affects the discovery potential, and it can be seen that the distributions are reduced and shifted towards higher masses and momenta.

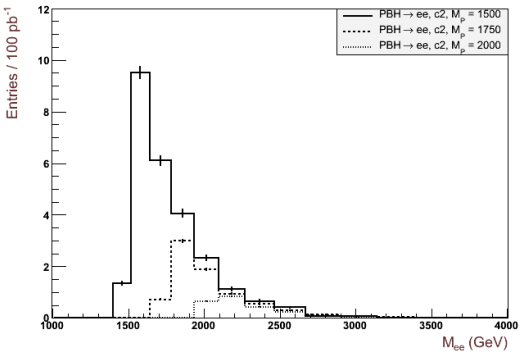
4.4 Alternate conservation laws

When flavours are not conserved, a multitude of decay channels are rendered possible, e.g. processes such as $ud \to bt$. Thus, heavy-flavour final states become much more frequent. Taking a quick glance at Table 3.2 again, it can be seen that the electron + muon final state is now allowed. However, the di-electron and di-muon branching ratios have decreased by $\sim 33 - 35\%$. The $Z + X$ ratio also decreases, but by a smaller amount, $\sim 17 - 18\%$. This can be seen in Figure 4.18, where invariant mass and p_T distributions for the same-flavour di-lepton channels have been plotted for various sets of conservation laws. With the statistics available, production of PBHs conserving B and L (c1) cannot be distinguished from production of PBHs conserving only B - L (c3). The lepton + jet decay channel will be well suited for this, but that is well beyond the scope of this thesis.

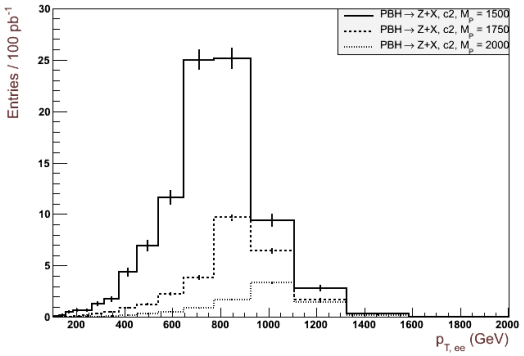
The mixed decay channel branching ratio increases by a factor ~ 10 when flavour-conservation is turned off, and thus it becomes an important source of signal in the sub-TeV region, see Figure 4.19. These events are mainly due to (semi-)leptonic W, tau, bottom and top decay, resulting in a sizeable fraction of the energy being carried away by hadrons and neutrinos (E_T^{miss}). The reconstructed di-lepton mass is thus smaller than the true PBH mass. Note that without conservation of flavour, the mixed decay channel contains single prompt lepton events as well, e.g. electron + tau. The rather large contribution from the $X + X$ decay channel in the di-muon channel is due to an increased occurrence of the relatively heavy charm quark as well as bottom or top quarks accompanied by a lighter quark.



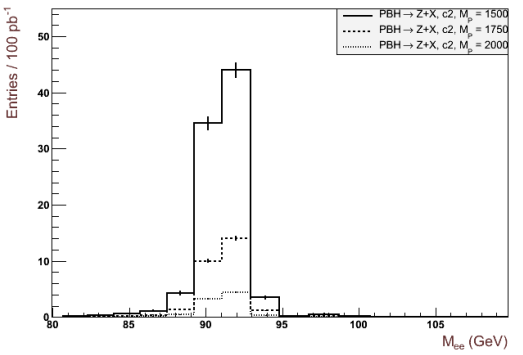
(a) PBH $\rightarrow ee$, electron p_T



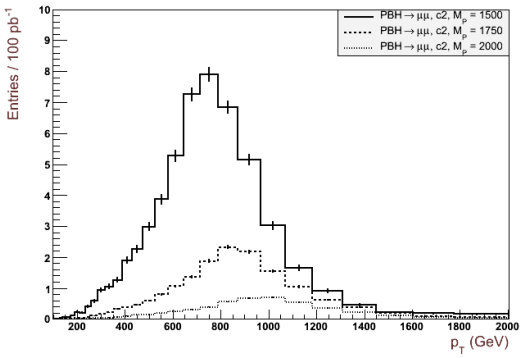
(b) PBH $\rightarrow ee$, invariant mass



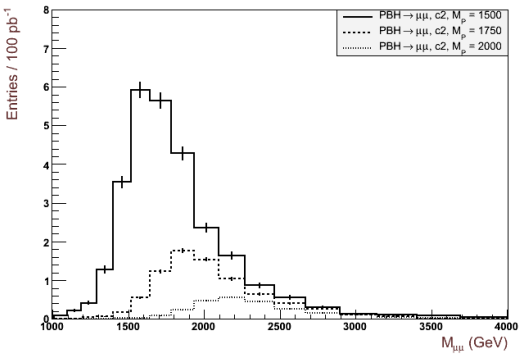
(c) PBH $\rightarrow Z + X \rightarrow ee + X$, di-electron p_T



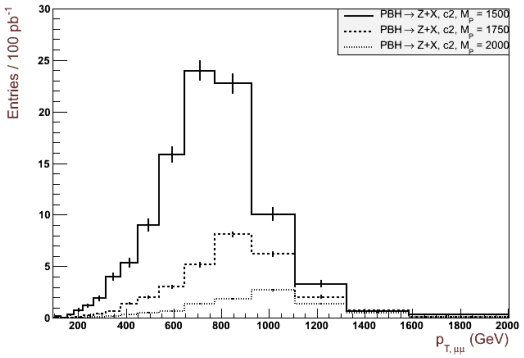
(d) PBH $\rightarrow Z + X \rightarrow ee + X$, invariant mass



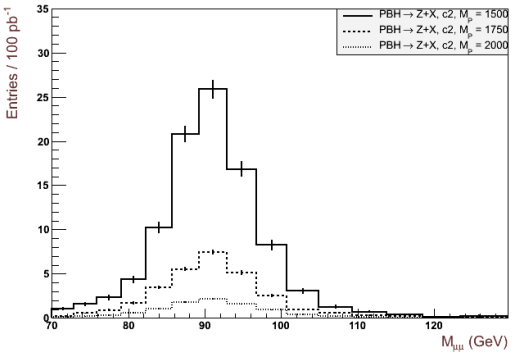
(e) PBH $\rightarrow \mu\mu$, muon p_T



(f) PBH $\rightarrow \mu\mu$, invariant mass

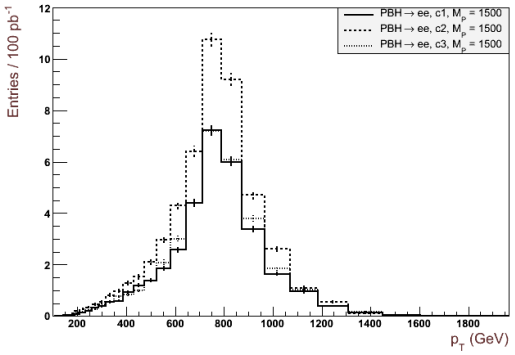


(g) PBH $\rightarrow Z + X$, di-muon p_T

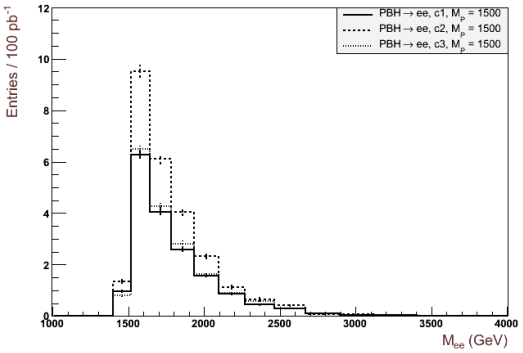


(h) PBH $\rightarrow Z + X$, invariant mass

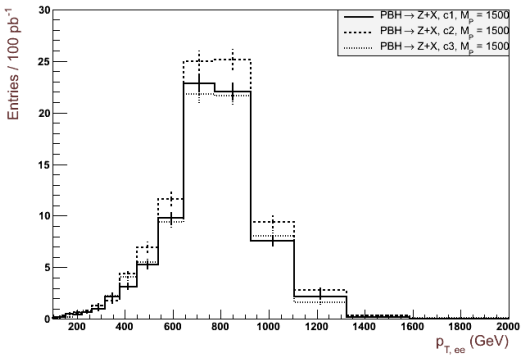
Figure 4.17: Invariant mass and p_T distributions for different Planck masses in the same-flavour di-lepton channels.



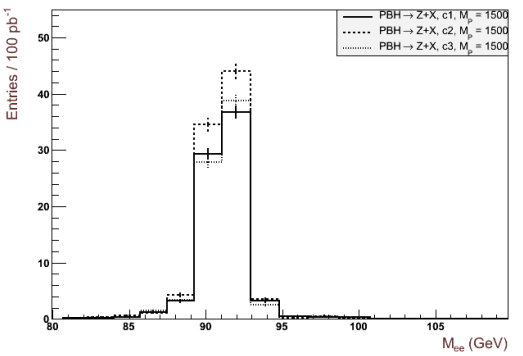
(a) PBH $\rightarrow ee$, electron p_T



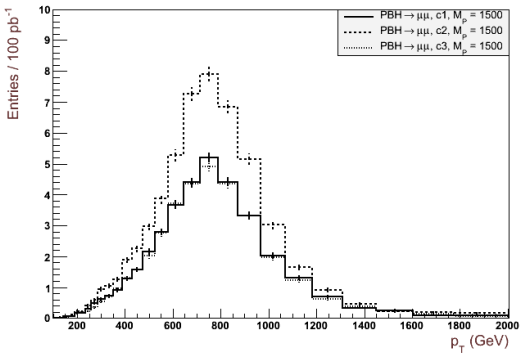
(b) PBH $\rightarrow ee$, invariant mass



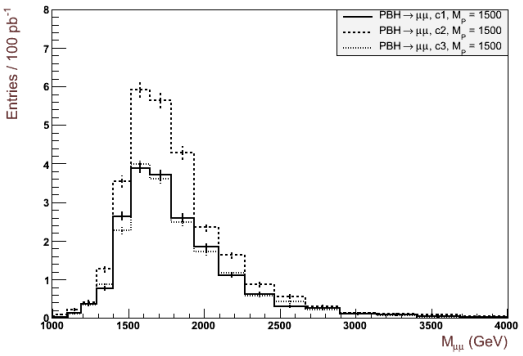
(c) PBH $\rightarrow Z + X \rightarrow ee + X$, di-electron p_T



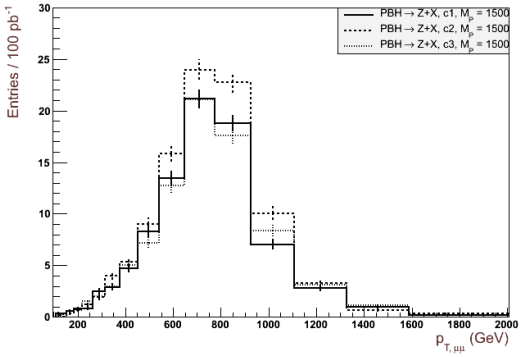
(d) PBH $\rightarrow Z + X \rightarrow ee + X$, invariant mass



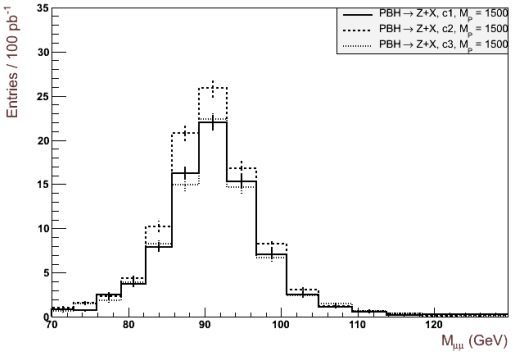
(e) PBH $\rightarrow \mu\mu$, muon p_T



(f) PBH $\rightarrow \mu\mu$, invariant mass

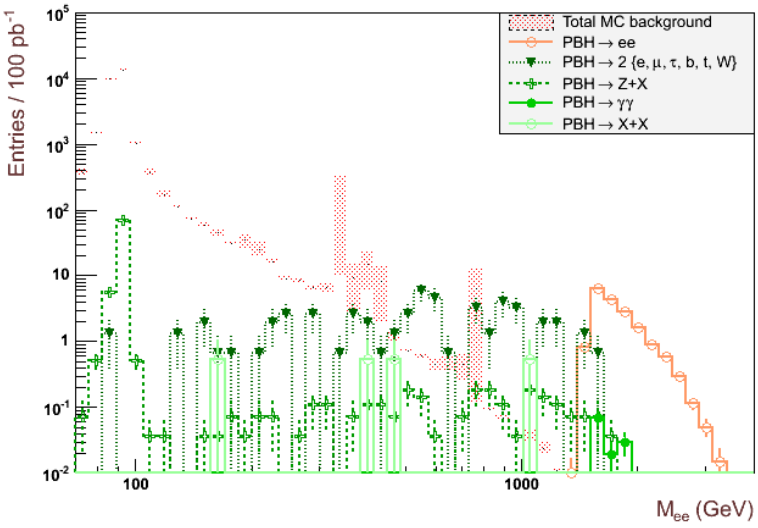


(g) PBH $\rightarrow Z + X$, di-muon p_T

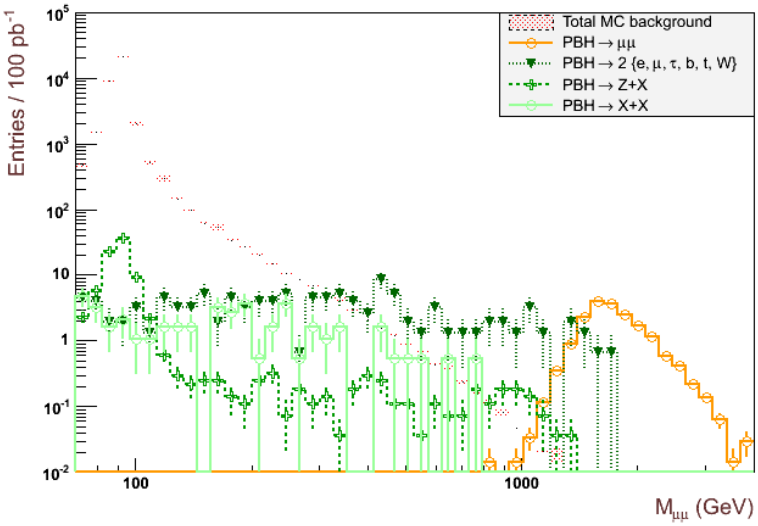


(h) PBH $\rightarrow Z + X$, invariant mass

Figure 4.18: Invariant mass and p_T distributions for different sets of conservation laws in the same-flavour di-lepton channels (continued).

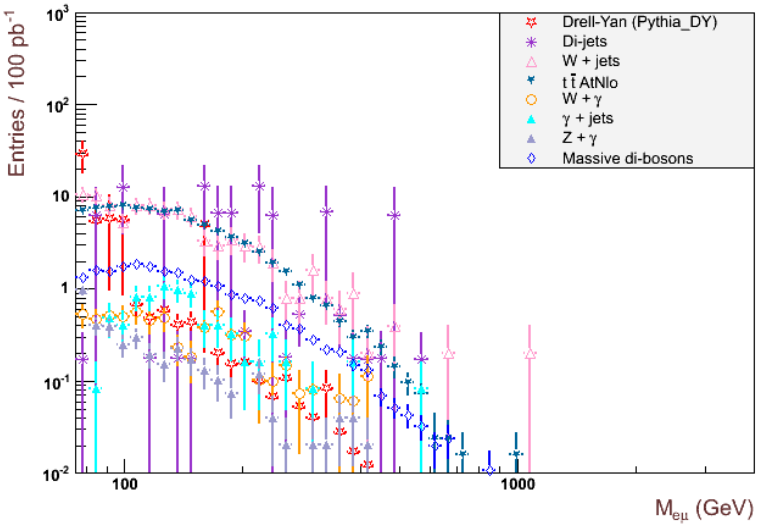


(a) Di-electron signal + SM background

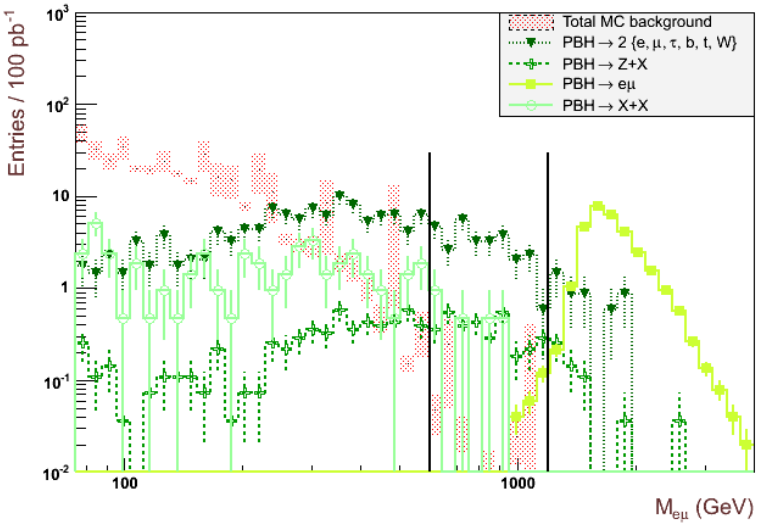


(b) Di-muon Signal + SM background

Figure 4.19: Invariant mass distribution of high-mass same-flavour di-leptons. Here, the signal is PBHs which only conserve B - L (c1).



(a) Electron + muon SM background components



(b) Electron + muon signal + SM background

Figure 4.20: Invariant mass distribution of electron + muon pairs, where the lines in (a) mark 600 and 1200 GeV. The background in (b) is the sum of the background in (a), and the signal is PBHs which only conserve B - L (c1).

4.4.1 Opposite-flavour di-leptons

For opposite-flavour di-leptons (only the $e\mu$ final state is considered) the following event selection is used:

Event

- Same primary vertex cuts as for the di-muon channel
- Muon or electron trigger must be fired.

Muon

- Same selection as for di-muons (no isolation or hits cuts)

Electron

- Same selection as for di-electrons.

Di-lepton

- Invariant mass > 60 GeV
- No requirement on opposite charge

This particular channel has very little SM background, with the important contributions being di-jets, di-taus (Drell-Yan), W + jets and $t\bar{t}$ production. As can be seen in Figure 4.20, signal from the mixed and electron + muon decay channels dominate for invariant masses above $200-300$ GeV. Placing a cut at 600 or 1200 GeV yields a signal-to-background ratio of 115 ± 56 and 25000 ± 11000, respectively. The expected number of events after each of the two cuts are given in Tables 4.12 and 4.13.

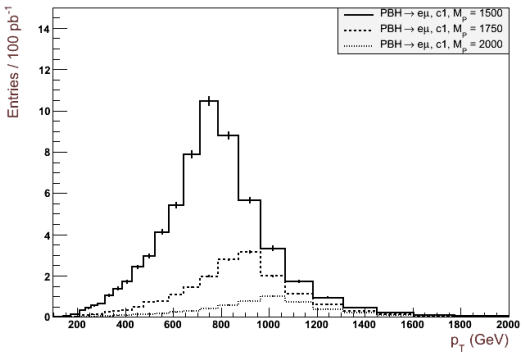
Dataset	events / pb^{-1}	$\pm$ stat. uncertainty
Drell-Yan	0.0101	$\pm$0.0024
W + jet	0.40	$\pm$0.28
Di-jets	0.0126	$\pm$0.0056
Top	0.088	$\pm$0.033
Massive di-bosons	0.074	$\pm$0.016
γ + jet	0.0070	$\pm$0.0014
PBH $\rightarrow ee$	0.0196	$\pm$0.0098
PBH $\rightarrow \mu\mu$	0.024	$\pm$0.011
PBH $\rightarrow Z + X$	3.72	$\pm$0.37
PBH $\rightarrow 2\{e, \mu, \tau, b, t, W\}$	31.5	$\pm$3.0
PBH $\rightarrow \gamma\gamma$	0.0051	$\pm$0.0051
PBH $\rightarrow e\mu$	30.48	$\pm$0.24
PBH $\rightarrow X + X$	2.3	$\pm$1.0

Table 4.12: Number of expected electron + muon events with an invariant mass above 600 GeV. The signal model is $M_P = 1.5$ TeV and conservation of B and L (c1).

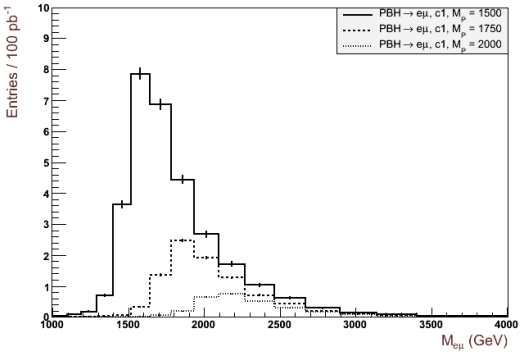
In Figure 4.21 the leading lepton p_T and invariant mass distributions for the PBH$\rightarrow e\mu$ samples have been plotted for different Planck masses and the two relevant sets of conservation laws. One can see that $e\mu$-pairs of PBHs conserving B and L (c1) can hardly be distinguished from those of PBHs which only conserve B - L (c3).

Dataset	events / pb^{-1}	± stat. uncertainty
Drell-Yan	0.0001657	±0.0000098
Di-jets	0.000024	±0.000017
Massive di-bosons	0.00039	±0.00039
γ + jet	0.00084	±0.00048
PBH $\rightarrow \mu\mu$	0.0098	±0.0069
PBH $\rightarrow Z + X$	0.58	±0.14
PBH $\rightarrow 2\{e, \mu, \tau, b, t, W\}$	4.6	±1.2
PBH $\rightarrow e\mu$	30.27	±0.24

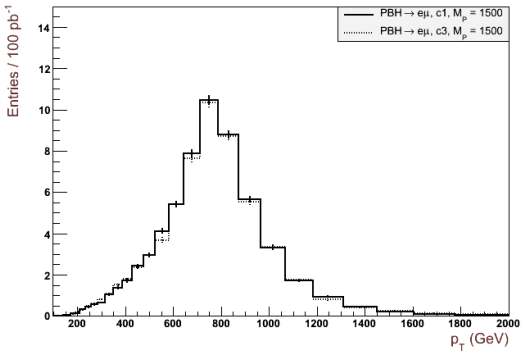
Table 4.13: Number of expected electron + muon events with an invariant mass above 1200 GeV. The signal model is $M_P = 1.5$ TeV and flavour conservation (c2).



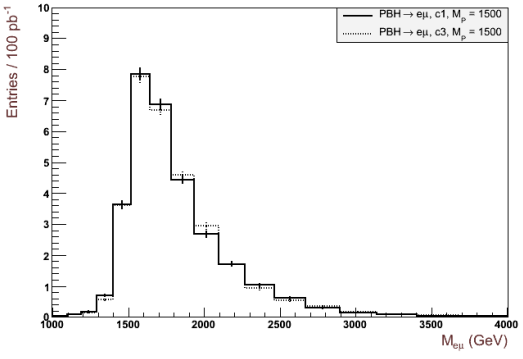
(a) PBH $\rightarrow e\mu$, lepton p_T



(b) PBH $\rightarrow e\mu$, invariant mass



(c) PBH $\rightarrow e\mu$, lepton p_T



(d) PBH $\rightarrow e\mu$, invariant mass

Figure 4.21: Invariant mass and p_T distributions for different Planck masses and sets of conservation laws in the opposite-flavour di-lepton channel.

The first search for a heavy particle decaying into an electron + muon final state with the ATLAS detector in $\sqrt{s} = 7$ TeV pp collisions at the LHC is presented in [66]. With the data recorded by the ATLAS detector during 2010, no excess above the SM background expectation was observed. As a result, exclusions at 95% confidence level were placed on two models. Tau sneutrinos with a mass below 0.75 TeV in an R-parity violating supersymmetric model, as well as lepton-flavour violating Z'-like vector bosons with masses of 0.70 to 1.00 TeV, were excluded.

4.5 The statistics of discovery and exclusion

The results in the previous sections will now be used to estimate how well the various PBH models can be seen in the ATLAS data, or conversely, which models may be excluded based on the lack of evidence in the data. Although the statistics of discovery and exclusion have advanced a lot in the last 15 years, it is still not a textbook subject and worth reviewing briefly before computing potential discovery significances and exclusion limits.

Compared to the common QCD processes that have cross sections as large as ~ 1 mb, PBH production at the LHC is expected to be rare[1]. It is also reasonable to believe that the cross sections are constant in time, provided that the center of mass energy does not change and that each event is independent of all other events. PBH production can thus be considered a Poisson process, a collection of random variables $N(t) : t \geq 0$, where $N(t)$ is the number of events observed in the time interval $[0, t]$. The probability to observe k events is then given by

$$P(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}, \tag{4.10}$$

where λ is the expected number of events,

$$\lambda = \sigma\varepsilon \int_0^t L dt'. \tag{4.11}$$

Here, σ is the cross section, ε the efficiency and L the luminosity.

In a counting experiment, events satisfying certain criteria are counted. If the number of observed events n_{obs} is significantly greater than the expected number of background events b, the background hypothesis is rejected. The statistical significance S is commonly expressed as a number of standard deviations of a normal distribution and is thus defined by

$$p = \int_S^\infty \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx, \tag{4.12}$$

where p is the p-value, which in this context is interpreted as the probability to falsely make a discovery in the presence of background only. The general agreement is that discovery can be claimed if the significance is greater than five, i.e. a 5σ deviation. This corresponds to a p-value

$$p = 2.87 \times 10^{-7},$$

a small number indeed.

On the other hand, if n_{obs} is significantly less than the number of expected signal and background events $s + b$, the signal + background hypothesis is rejected at a confidence level

[1] As discussed in Section 1.3, hard perturbative processes would be highly suppressed for $\sqrt{s} \gg M_P$. In this regime, thermal black holes would be produced, and at some point, thermal BH production would surpass QCD processes.

$1-p$. For a counting experiment, the p-value CL_{s+b} is defined as the probability to observe the number of events observed in the experiment or less given the signal + background hypothesis,

$$p = CL_{s+b} = P\left(n \leq n_{obs} | s+b\right). \tag{4.13}$$

One usually excludes at a 95% confidence level, corresponding to a p-value of 5%. The CL_{s+b} exclusion is a statement about the signal + background hypothesis and not the signal hypothesis, and this may lead to strange results. If the observation in an experiment falls significantly below the background expectation, one would question whether the background is fully understood. Still, by following the CL_{s+b} method, this experiment would quote a stronger limit on the signal hypothesis than an identical experiment in which the background is properly understood and not "conservatively" over-estimated.

A method which protects against overly strong limits, at the cost of some over-coverage[2], is the CL_s approach [67]. Here one attempts to make a statement about the signal hypothesis instead. The "p-value" CL_s is defined as the p-value for the signal + background hypothesis weighted by a factor $1/CL_b$,

$$CL_s = \frac{CL_{s+b}}{CL_b}, \tag{4.14}$$

where, analogously to CL_{s+b}, CL_b is defined as the probability to observe the number of events observed in the experiment or less given the background only hypothesis,

$$CL_b = P\left(n \leq n_{obs} | b\right). \tag{4.15}$$

The signal hypothesis is considered to be excluded at a confidence level $1 - CL_s$.

The recently proposed *power-constrained* procedure [68] is another method which protects against excluding models to which one has little or no sensitivity. The sensitivity is measured from the p-value of the signal + background hypothesis for a sample of background-only pseudo-experiments. One simply does not allow the observed limit on a given model at 95% confidence level to fluctuate below the -1σ expected limit. This is the new ATLAS standard for setting exclusion limits and will be utilized in the next chapter.

4.5.1 The likelihood ratio

There are many methods for estimating the significance, and some of the frequently used ones were compared in [69]. A method based on the likelihood ratio of the signal + background and the background only hypotheses was found to give the best performance. The likelihood is the probability of the data x given a model $f(\theta)$. If each observation x_i is statistically independent of the others, the likelihood function can be expressed as a multiplication of p.d.f.s for individual observations,

$$f(x_1, \ldots, x_n | \theta) = \prod_{i=1}^{n} f(x_i | \theta). \tag{4.16}$$

For a Poisson process, the likelihood ratio of the signal + background and the background only hypotheses takes the form

$$Q = \frac{\mathcal{L}_{b+s}}{\mathcal{L}_b} = \frac{P(n_{obs}|s+b)}{P(n_{obs}|b)} = \left(1 + \frac{s}{b}\right)^n e^{-s}. \tag{4.17}$$

In an experiment with several statistically independent channels, the likelihood ratios can easily be combined,

$$Q = Q_1 \times Q_2 \times \cdots \times Q_n. \tag{4.18}$$

[2]Over-coverage means that the signal hypothesis would be falsely excluded less frequently than one minus the quoted confidence level.

The indices $1, 2, \ldots, n$ can be bins in a single search channel as well as different channels. Wilks' theorem states that when the sample size approaches infinity, the test statistic $-2\ln Q$ for a nested model will be asymptotically χ^2-distributed. Thus, the χ^2 corresponding to the desired statistical significance can be compared to $-2\ln Q$ and used as an approximate statistical test. For a Poisson process, the significance is thus estimated by

$$S_{L2} = \sqrt{2\ln Q} = \sqrt{2\left[n\ln\left(1+\frac{s}{b}\right) - s\right]}. \tag{4.19}$$

This is further used to approximate the p-value for the signal + background hypothesis,

$$CL_{s+b} \approx \frac{1}{2}P\left(\chi^2 \leq S_{L2}^2\right) = \frac{1}{2}P\left(\chi^2 \leq 2\ln Q\right), \tag{4.20}$$

where the division by two is due to the fact that a one-sided hypothesis test is being performed.

The Neyman-Pearson lemma [70] states that the likelihood ratio test-statistic maximizes the probability to both discover a true signal and exclude a false signal hypothesis. This is illustrated in [71]. In the absence of systematic uncertainties, the optimal search region is in fact the one including 100% of the signal, provided that a discriminating variable enters the likelihood ratio. A region with poor signal-to-background ratio will at worst enter the likelihood ratio as a factor 1. However, when uncertainties come into play, the expected improvement of accepting certain parts of the signal region must be weighed against the effects of the additional uncertainties.

4.5.2 Expectations for discovery and exclusion potential

A multi-bin search will be done for each of the search channels. The estimated expected significances are calculated by inserting $s+b$ for n in Eq. 4.19[3], while the expected CL_{s+b} exclusion limits are estimated by inserting b for n instead. From Eq. 4.19 it is clear that bins with zero expected background and a non-zero signal yield an infinite significance, and so this must be avoided. Moreover, one can only observe zero or a positive integer number of events in a real experiment. To account for this fact and to avoid bins with zero background, which is definitively a problem in the TeV region, a simple algorithm was used.

- Start with the last bin (make sure there are neither signal nor background entries above).
- If $b \leq 0$[4] OR $n < 1$, add another bin.
- When the accumulated region yields at least one expected event n and a positive number of expected background events:
  - Calculate s/b and take the floor of n, $\lfloor n \rfloor$.
  - Store the remaining signal s' and background b' in temporary variables.

$$s' = \left(\frac{s}{n}\right)\left[n - \lfloor n \rfloor\right] \tag{4.21}$$

$$b' = \left(\frac{b}{n}\right)\left[n - \lfloor n \rfloor\right] \tag{4.22}$$

  - Add the log likelihood of $\lfloor n \rfloor$, $s - s'$ and $b - b'$ to the log likelihood sum.
- Start with $s = s'$ and $b = b'$ and repeat for the bin below the last one added.

[3] This is the so-called "Asimov dataset" [72].

[4] The events of the T1_McAtNlo sample are weighted. Thus, some bins may contain negative background.

It should be noted that the algorithm is applied assuming an integrated luminosity of 100 pb^{-1}. In order to estimate the luminosity at which e.g. a 5σ excess or more w.r.t. background is expected, the results will be scaled accordingly. Hence, only the significances found for 100 pb^{-1} account for the fact that one can only observe an integer number of events in search channel and bin.

Two statistically independent search regions are defined for the same-flavour di-lepton channels. The boosted Z region is defined as $M_{ll} \in (60, 120)$ GeV, approximately symmetric around the Z mass, while the high-mass region is defined as $M_{ll} > 120$ GeV. Note that boosted Zs are searched for in the di-lepton p_T spectrum. The search region for opposite-flavour di-leptons is defined as $M_{e\mu} > 75$ GeV, the coverage of the binned Drell-Yan samples. All background and signal samples were considered, and the bin width was set to 10 GeV in both the invariant mass and di-lepton p_T spectra.

Using the above-mentioned algorithm and search regions, one can see from Table 4.14 that a 5σ excess or more w.r.t. background is expected at very low luminosities, i.e. in the order of a few to tens of pb^{-1}. However, it is customary to not claim discovery with less than ten signal events. Thus, the luminosities given in Table 4.15 are better indicatives of the discovery potential. It is also worth noting that a coarser binning slightly reduces the sensitivity. This is seen by comparing Tables 4.16 and 4.17, where the expected significances for 100 pb^{-1} using bin widths of 10 GeV and 100 GeV, respectively, are listed. If no discovery is made, it is expected that all signal models considered can be excluded at a 95% confidence level at rather low luminosities, see Table 4.18.

Shortly summarized, if one ignores systematic uncertainties, the expected significances for the considered signal models are quite large. With an integrated luminosity of ~ 40 pb^{-1} in the 2010 data, it is fair to believe that several of the considered signal models can either be discovered or excluded, even when taking systematic uncertainties into account (which of course reduce the sensitivity). The issue of systematic uncertainties will be addressed in the next chapter.

M_P (GeV)	1500			1750			2000		
Conservation law set	c1	c2	c3	c1	c2	c3	c1	c2	c3
High-mass ee	7.93	6.83	6.06	27.6	23.1	21.8	104	91.3	71.1
High-mass $\mu\mu$	6.09	6.90	5.40	24.5	25.9	19.5	88.5	98.0	74.1
High-mass $e\mu$	1.78	150	1.41	6.29	805	4.86	22.3	6810	19.7
Boosted Z ee	0.871	0.744	0.896	3.28	2.63	3.25	13.3	10.1	12.1
Boosted Z $\mu\mu$	1.12	1.06	1.23	4.80	4.81	4.78	17.7	19.8	16.6

Table 4.14: Luminosity (pb^{-1}) at which a 5σ excess or more w.r.t. background is expected. The minimal bin width is set to 10 GeV.

M_P (GeV)	1500			1750			2000		
Conservation law set	c1	c2	c3	c1	c2	c3	c1	c2	c3
High-mass ee	17.0	30.7	13.5	47.3	101	37.9	154	350	116
High-mass $\mu\mu$	7.44	24.7	6.81	24.5	83.9	19.5	88.5	256	74.1
High-mass $e\mu$	4.27	150	3.83	11.9	805	9.65	29.1	6810	28.4
Boosted Z ee	12.6	11.0	12.8	39.4	36.2	39.1	123	113	123
Boosted Z $\mu\mu$	7.81	9.87	7.58	23.2	32.1	23.2	65.2	98.9	66.4

Table 4.15: Luminosity (pb^{-1}) at which at least 10 signal events and a 5σ excess or more w.r.t. background is expected. The minimal bin width is set to 10 GeV.

M_P (GeV)	1500			1750			2000		
Conservation law set	c1	c2	c3	c1	c2	c3	c1	c2	c3
High-mass ee	17.8	19.1	20.3	9.52	10.4	10.7	4.90	5.23	5.93
High-mass $\mu\mu$	20.3	19.0	21.5	10.1	9.82	11.3	5.32	5.05	5.81
High-mass $e\mu$	37.5	4.08	42.1	19.9	1.76	22.7	10.6	0.606	11.3
Boosted Z ee	53.6	58.0	52.8	27.6	30.8	27.7	13.7	15.7	14.4
Boosted Z $\mu\mu$	47.2	48.6	45.1	22.8	22.8	22.9	11.9	11.2	12.3

Table 4.16: Expected significance for 100 pb^{-1} of data with the bin width set to 10 GeV.

M_P (GeV)	1500			1750			2000		
Conservation law set	c1	c2	c3	c1	c2	c3	c1	c2	c3
High-mass ee	17.0	18.7	19.5	9.25	10.3	10.4	4.80	5.23	5.73
High-mass $\mu\mu$	20.0	18.9	20.8	9.96	9.78	11.1	5.22	4.98	5.68
High-mass $e\mu$	35.1	3.66	38.7	18.3	1.68	20.9	9.96	0.598	9.86
Boosted Z ee	39.3	43.5	39.7	22.3	24.9	21.7	11.8	12.8	12.0
Boosted Z $\mu\mu$	40.0	39.8	39.4	21.5	20.9	21.7	11.5	11.1	11.9

Table 4.17: Expected significance for 100 pb^{-1} of data with the minimal bin width set to 100 GeV.

M_P (GeV)	1500			1750			2000		
Conservation law set	c1	c2	c3	c1	c2	c3	c1	c2	c3
High-mass ee	10.6	16.2	7.42	45.3	88.3	30.6	277	694	176
High-mass $\mu\mu$	4.69	10.1	4.06	19.6	41.6	13.2	65.3	186	49.1
High-mass $e\mu$	2.02	59.6	1.68	6.51	246	5.05	19.3	1560	19.0
Boosted Z ee	3.92	3.36	3.98	13.9	12.4	13.9	55.2	48.1	53.2
Boosted Z $\mu\mu$	2.57	2.99	2.66	8.28	10.8	8.62	28.8	40.5	27.0

Table 4.18: Luminosity (pb^{-1}) at which a 95% confidence level exclusion of the signal + background hypothesis is expected. The minimal bin width is set to 10 GeV.

Chapter 5

2010 data analysis

In the following, official 2010 data and MC background of the Release 16 production, also called the "autumn reprocessing", will be analyzed. The signal simulation samples were privately produced by the author of this thesis, but fully simulated and reconstructed with the same detector conditions and options as the official samples. A complete list of MC datasets with cross sections and integrated luminosities used in this analysis is given in Appendix C.2.

5.1 Pre-selection

In the following analysis, energy correction and smearing procedures will be applied to both data and MC. A lepton with p_T just below 25 GeV can then have its p_T corrected or smeared to just above 25 GeV. Hence, the pre-selection from Section 4.1.3 is loosened:

Electrons

- Cluster $p_T > 20$ GeV
- $|\eta| < 2.5$
- Author 1 or 3
- Loose

Muons

- Combined $p_T > 20$ GeV
- CombinedMuon
- Combined $|\eta| < 2.5$

Event

- If number of remaining leptons ≥ 2, accept event

5.2 Signal simulation

In order to obtain more reliable results, a full simulation of PBHs for $M_P = 2.0$ TeV[1] was done for all three sets of conservation laws. The event-filtering in BlackMax was extended to a total of eight classes of decay channels. Production of a lepton accompanied by a tau-lepton or heavy quark is a rather important channel when flavour, B and L are all violated and was

[1]Due to lack of time, only the single amss point $M_P = 2.0$ TeV was simulated. A very rough extrapolation to nearby Planck masses is described in Section 5.13.

thus granted its own class. This class also contains lepton + light quark events, which were previously contained in the $X + X$ class. The eight decay channel classes are listed in Table 5.1 along with branching ratios for $M_P = 2.0$ TeV. The PBH production cross section here is 1.257×10^2 pb with an uncertainty of 0.6%.

N conservation laws	1 (c3)	3 (c1)	12 (c2)
$Z + X$	2.41	2.41	2.92
e^+e^-	0.018	0.018	0.027
$\mu^+\mu^-$	0.018	0.018	0.027
$\gamma\gamma$	0.0091	0.0091	0.014
$2 \times \{\tau, b, t, W\}$	8.63	9.02	0.66
$e + \mu$	0.036	0.036	0
$e/\mu + X$	1.17	0.18	0
$X + X$	77.18	88.31	96.35

Table 5.1: PBH decay channels with branching ratios (%) for $M_P = 2.0$ TeV listed in the same order as they are tested for in BlackMax. For example, an electron + muon event will always fire on the $e + \mu$ test first and will never be stored as an $e/\mu + X$ event.

5.2.1 Transform options

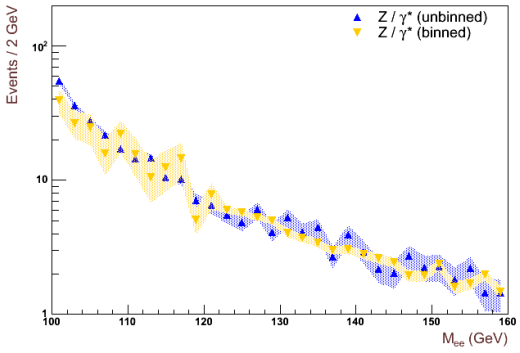
The signal dataset were simulated and reconstructed with the same transform options as an official BlackMax dataset (run number 105453) of the 2010 Release 16 production. These options correspond to event generation tag e660, simulation tag s933 and digitization and reconstruction tag r1831. The private signal and official background datasets thus have very similar tags and are consistent with respect to detector conditions. Each signal event was required to contain at least one true lepton after hadronization, prompt decays and underlying event modeling with Pythia. The digitization approach corresponding to tag r1831 accounts for the effects of the numerous elastic collisions (minimum-bias) and low-energy photons and neutrons generated from interactions in walls and shielding (cavern). The input minimum-bias events are used to form pile-up (see Section 5.5), which along with cavern background events are superimposed on top of the physics events. The options given for each transform are listed in in Appendix D.2.

5.3 Binned and unbinned background samples

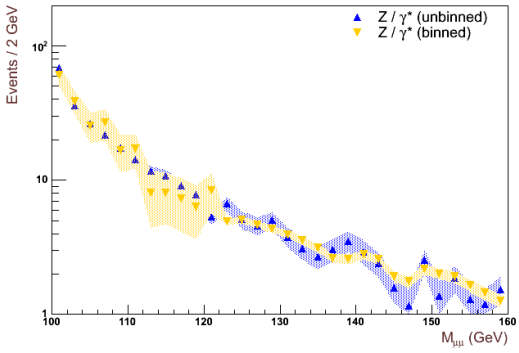
Two types of samples are used to account for the Drell-Yan background (see Table 3.5), that is, binned in mass windows (a set of datasets each covering a small mass window) and unbinned (a large dataset covering the full mass range). From Figure 5.1 one can see that the unbinned sample has better statistics in the mass-range below 120 GeV, i.e. in the vicinity of the Z-peak, while the binned samples have better statistics in the mass-range above. In the following analysis, events from the unbinned sample are accepted if the generator level mass is ≤ 120 GeV, while events from the binned samples are accepted if the generator level mass is > 120 GeV.

5.4 Cross sections

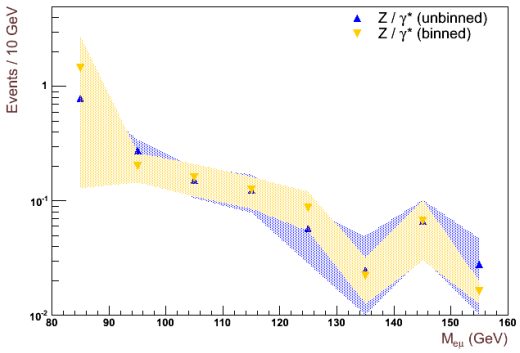
In the 2010 ATLAS Monte Carlo production, Drell-Yan processes were simulated using the leading-order (LO) generator Pythia and MRST2007LO* PDF. The production of lepton-pairs via photon and Z boson exchange has been calculated up to next-to-next-to-leading-order



(a) Di-electron mass



(b) Di-muon mass



(c) Electron + muon mass

Figure 5.1: Invariant mass distribution with statistical uncertainty bands for the binned (yellow) and unbinned (blue) Drell-Yan samples; (a) di-electron samples, (b) di-muon samples and (c) di-tau samples.

(NNLO) in QCD using the MSTW2008NLLO PDF [73]. The cross section-ratios (K-factors) have a strong mass-dependence and are defined as

$$K^{*}_{NNLO}(M_{ll}) = \frac{d\sigma_{NNLO}}{dM^2_{ll}}(MSTW2008NNLO) \Big/ \frac{d\sigma_{LO}}{dM^2_{ll}}(MSTW2007LO^{*}).$$

In Table 5.2 K-factors with symmetric and asymmetric uncertainties for several masses are listed.

In this analysis, cubic spline interpolation between the K-factors given in column five and the true (generator level) mass of the exchanged photon or Z is used to evaluate the event weight. The highest Drell-Yan mass given in the table is 3000 GeV. The few simulated events with a larger mass will simply be weighted by the K-factor corresponding to 3000 GeV.

The W + jets cross section has been calculated to NNLO precision, whereas the massive di-boson cross section has been calculated to NLO precision. Regarding the W + jets cross section, the recommendation of [75] is to use the central values from QCD NNLO calculations by FEWZ [76] and quote a conservative 5% theoretical uncertainty. By applying NNLO corrections, the W + jets cross section increases by ∼26%. For massive di-bosons, the recommendation is to quote NLO calculations from MCFM [77] with MSTW2008NLO PDF as the reference cross section. A total theoretical uncertainty of 5% for WW and ZZ and 7% for WZ is suggested. From approximate NNLO calculations, the cross section for top-pair production is normalized to 164.6 pb [78]. The K-factors with uncertainties are listed in Table 5.3.

5.5 Pile-up simulation

At high luminosities, additional soft interactions occur in the same bunch crossing as the hard scatter event. This effect is called *in-time pile-up*. Starting with run 165591, the LHC was running with bunch trains with a mere 150 ns bunch separation. The small separation causes so-called *out-of-time* pile-up in addition, e.g. calorimeter signals tend not to return to zero until several of the closely spaced bunches have crossed. MC samples with the r1831 or r1833 tag were reconstructed with both in-time and out-of-time pile-up, and only these samples will be used in the following. The in-time pile-up is simulated with an average of 2.2 in-time overlaid interactions per event, while the out-of-time pile-up is simulated with double trains separated by 225 ns. Each train contains 8 filled bunches with a 150 ns separation.

According to [79] the best estimator of the number of additional interactions in one beam crossing is the Poisson-distributed number of reconstructed primary vertices. While this observable varies in data, it is constant in MC. In order for the multiplicity distribution of vertices in MC to match that of data, re-weighting is necessary. The event weights used are listed in Table 5.4.

5.6 Collision data

It is important that the relevant parts of the detector were operational during the data taking. A collection of subsequent events in which the detector status remains the same is called a *luminosity block*. In order to filter out data which are declared flawed or bad for given physics objects, e.g. electrons or muons, so-called *Good Run Lists* (GRLs) are implemented in the analysis. These lists contain all runs and luminosity blocks fulfilling a set of requirements. Since electrons and muons are dependent on different sub-systems, one GRL is needed for each analysis. More details on the data quality flags and good run lists can be found in [80].

$M_{\ell\ell}$ [GeV]	$M_{\ell\ell}^2 \frac{d\sigma_{NNLO}}{dM_{\ell\ell}^2}$ [nb] MSTW 2008NNLO	$\frac{\sigma_{NLO}}{\sigma_{LO}}$ MSTW 2008NLO/LO	$\frac{\sigma_{NNLO}}{\sigma_{LO}}$ MSTW 2008NNLO/LO	$\frac{\sigma_{NNLO}}{\sigma_{LO}}$ MSTW/MRST 2008NNLO/2007LO*	Δ_r [%] PDF uncert. 68% C.L.	Δ_r^+ [%]	Δ_r^- [%]	Δ_r [%] PDF uncert. 90% C.L.	Δ_r^+ [%]	Δ_r^- [%]
10	0.465E+01	1.022	1.081	1.166	5.5	7.5	-4.5	9.3	12.2	-8.8
20	0.917E+00	1.088	1.133	1.138	2.3	3.1	-2.0	4.5	5.5	-4.3
30	0.327E+00	1.124	1.166	1.142	1.8	2.1	-1.8	3.7	4.0	-3.7
40	0.154E+00	1.150	1.191	1.146	1.7	1.8	-1.7	3.5	3.6	-3.5
50	0.884E-01	1.171	1.212	1.149	1.7	1.7	-1.7	3.4	3.5	-3.3
60	0.633E-01	1.190	1.231	1.148	1.7	1.7	-1.6	3.3	3.5	-3.2
70	0.665E-01	1.208	1.248	1.144	1.7	1.7	-1.6	3.3	3.5	-3.2
80	0.155E+00	1.225	1.262	1.138	1.7	1.7	-1.6	3.3	3.5	-3.2
91.12	0.113E+02	1.239	1.275	1.136	1.6	1.7	-1.6	3.3	3.5	-3.1
100	0.236E+00	1.246	1.282	1.138	1.6	1.7	-1.6	3.2	3.5	-3.1
125	0.207E-01	1.263	1.299	1.145	1.6	1.7	-1.5	3.2	3.5	-3.0
150	0.784E-02	1.277	1.312	1.149	1.6	1.7	-1.5	3.1	3.5	-3.0
175	0.405E-02	1.287	1.323	1.151	1.6	1.7	-1.5	3.2	3.5	-2.9
200	0.239E-02	1.296	1.331	1.151	1.6	1.7	-1.5	3.2	3.6	-2.9
250	0.104E-02	1.308	1.342	1.149	1.6	1.8	-1.4	3.3	3.7	-3.0
300	0.528E-03	1.316	1.349	1.146	1.6	1.8	-1.5	3.4	3.9	-3.1
400	0.179E-03	1.322	1.354	1.139	1.7	2.0	-1.5	3.6	4.2	-3.3
500	0.750E-04	1.321	1.352	1.131	1.8	2.1	-1.6	3.9	4.5	-3.5
600	0.357E-04	1.316	1.347	1.123	1.9	2.3	-1.7	4.1	4.8	-3.7
700	0.185E-04	1.310	1.339	1.114	2.0	2.5	-1.7	4.4	5.1	-3.8
800	0.101E-04	1.302	1.332	1.104	2.2	2.7	-1.8	4.6	5.5	-4.0
900	0.582E-05	1.295	1.324	1.093	2.3	2.9	-1.8	5.0	6.0	-4.1
1000	0.346E-05	1.288	1.316	1.080	2.5	3.2	-1.9	5.4	6.6	-4.3
1250	0.105E-05	1.271	1.300	1.041	3.2	4.1	-2.3	6.7	8.4	-5.3
1500	0.353E-06	1.257	1.290	0.990	4.2	5.4	-3.2	8.8	11.0	-7.0
1750	0.127E-06	1.247	1.286	0.929	5.5	7.1	-4.4	11.6	14.5	-9.3
2000	0.473E-07	1.241	1.288	0.860	7.3	9.2	-6.1	15.3	19.1	-12.3
2500	0.687E-08	1.230	1.300	0.712	11.9	14.5	-10.7	24.8	30.8	-20.3
3000	0.949E-09	1.199	1.295	0.563	17.2	20.3	-16.2	35.4	43.7	-29.6

Table 5.2: NNLO Drell-Yan lepton-pair production cross section $M_{ll}^2 d\sigma_{NNLO}/dM_{ll}^2$ as function of dilepton mass M_{ll} calculated with PHOZPR [74] and the MSTW2008NNLO PDF set, cross section ratios (K-factors) based on the MSTW2008LO, NLO and NNLO and MRST2008* PDF sets, symmetric and asymmetric PDF uncertainties for the NNLO cross section at 68% and 90% C.L., respectively.

Process	K-factor	uncertainty
W + jets	1.22×1.030	$\pm 5\%$
WW	1.55	$\pm 5\%$
ZZ	1.42	$\pm 5\%$
WZ	1.74	$\pm 7\%$
$t\bar{t}$	1.14	$+6.9\%/-9.5\%$

Table 5.3: K-factors used for W + jets and massive di-bosons.

# vertices	weight
1	1.7955
2	1.2259
3	0.8834
4	0.6890
5	0.5686
6	0.4945
7	0.4511
8	0.4583
9	0.4513
>9	0.5809

Table 5.4: Event weights for MC with bunch-train pile-up (r1831).

5.6.1 Data quality flags for electrons

In the electron analysis, version 3 (DetStatus-v03-repro05-01) of the data quality flags set by the ElectronGamma (EGamma) performance group was used to build a GRL. For electrons the important parts of ATLAS are the solenoid, the inner detector and the electromagnetic calorimeter. Thus, the requirements are:

- ATLGL green: data quality information has been reviewed;
- ATLSOL green: solenoid on and stable;
- L1CTP green: no clock or data header problems;
- L1CAL green: level 1 calorimeter trigger;
- TRELE green:electron trigger operating normally;
- TRGAM green: photon trigger operating normally;
- cp_eg_electron_barrel: include barrel region of the detector for electrons;
- cp_eg_electron_endcap: include endcap region of the detector for electrons;
- IDVX green: ID vertexing quality;
- LUMI green: offline luminosity OK.

5.6.2 Data quality flags for muons

Since no calorimeter information is used in this muon analysis, only muon spectrometer and inner detector related data quality flags are required. The Z' GRL, which will be used in this analysis, imposes the following requirements on the lumi blocks:

- ATLGL green: data quality information has been reviewed;
- ATLSOL green: solenoid on and stable;
- ATLTOR green: toroid on and stable;
- L1MUE green: level 1 MS end-cap (TGC) trigger;

- L1MUB green: level 1 MS barrel (RPC) trigger;
- TRMUO green: muon trigger operating normally;
- MSTACO, MMUIDCB green: muon reconstruction algorithms;
- MDTBA, MDTBC, MDTEA, MDTEC, CSCEA, CSCEC, RPCBA, RPCBC, TGCEA, TGCEC green: muon system operating normally;
- PIX0, PIXB, PIXEA, PIXEC, SCTB, SCTEA, SCTEC, TRTB, TRTEA, TRTEC, IDGL, IDAL green: inner detector operating normally;
- LUMI green: offline luminosity OK.

5.6.3 Luminosity

Accurate determination of the luminosity is essential, and special sub-detectors and algorithms are used for this task. Different detectors and algorithms agree within ±2%, but the absolute luminosity calibrations suffer from a 11% systematic uncertainty[2]. Uncertainties in the bunch intensities are the dominating term, and this is highly correlated across all methods [82]. The luminosity uncertainty is only applied to MC samples which are not normalized to data.

5.7 Preparing for electron analysis

In the di-electron analysis, the standard $Z' \rightarrow ee$ selection for Release 16 reconstructed data and MC, given in Table 5.5, is used. What follows is a description of the cuts and corrections applied to electrons in data and MC. Data periods and triggers used, as well as the cut flow for data, are also given.

5.7.1 Energy and direction

Since the covariance matrices of the track fitting had not been validated on data yet, the EGamma group recommended not to use combined four-momentum [83] for datasets reprocessed in the autumn 2010. This recommendation will be followed throughout the analysis. In most cases, the transverse energy will be calculated from track θ and cluster energy. If the track quality is poor, that is, the number of hits in the Pixel and SCT is less than four, only cluster information is used. Invariant mass is, however, always calculated from track directions and cluster energy. Taking the direction and position (at closest point to the primary vertex) from the track gives a resolution on the di-electron mass better than 1% above 400 GeV. For all other purposes, cluster η will be used [79].

5.7.2 Electron quality

The isEM variable is an unsigned integer, and thus there are 32 bits which can have a particular meaning. For example, bit 2 represents the cuts applied on hadronic leakage, whereas bits 20-22 represent track matching. In the data and MC reprocessed in the autumn, the isEM bit masks are missing the $\Delta\eta$ track match bit. This can be fixed by using the Medium WithTrackMatch definition, which is equivalent to the RobustMedium identification used in Release 15 [84]. If bitwise AND on isEM and Medium WithTrackMatch returns zero, the electron is accepted. To suppress background from photon conversions, hits in the innermost layer (B-layer) of the Pixel detector are required if expected.

[2]It should be noted that the updated uncertainty on the luminosity is determined to be 3.4% [81], but this analysis was frozen before the author became aware of the update. If the updated uncertainty had been incorporated, the final results of this analysis would most likely have been improved.

Cut type	variable	see Section
Good run list		5.6.1
At least one vertex with more than two tracks	vx_n, vx_nTracks	
Trigger		5.7.4, 5.7.5
Cluster-based electrons (author 1 or 3)	el_author	3.4.1
$\|\eta\|$ less than 2.47 and excluding crack region	el_cl_eta	5.7.1
Both medium	el_isEM	3.4.1, 5.7.2
E_T larger than 25 GeV	el_cl_E, el_tracktheta, el_nSCTHits, el_nPixHits, el_cl_pt	5.7.1
OQ map	el_cl_eta, el_cl_phi, RunNumber	4.1.8, 5.7.3
BLayer	el_expectHit-InBLayer, el_nBLHits	5.7.2
Invariant mass ≥ 70	el_tracktheta, el_trackphi, el_cl_E	5.7.2

Table 5.5: The event selection for di-electrons (standard Z' selection for Release 16 reconstructed data and MC).

5.7.3 Object Quality maps

The Object Quality maps were discussed in section 4.1.8. In the Release 16 production, Monte Carlo samples (mc10) were produced with detector conditions corresponding to Map 161730. Thus, only Maps 161730-169136 are used. To ensure full consistency between the treatment of data and MC, one can use the latest Map (169136) for all data and MC. By using this approach, 0.4% of the data is lost. Another method, which recovers the lost data, is to apply a luminosity-weighted method. In this analysis, the luminosity-weighted method is used. For each map above and including 161730, the corresponding efficiency was evaluated for the MC. To find the total efficiency, the efficiencies were weighted by the integrated luminosity corresponding to each of the maps. Hence,

$$\varepsilon = \frac{1}{\mathcal{L}} \sum_{Map} \mathcal{L}_{Map} \varepsilon_{Map}.$$

The list of maps, corresponding runs and luminosities are given in Table 5.6.

The final states of the various background processes may have different spatial distributions, and thus it is not very accurate to calculate an overall efficiency for all MC. For example, if there are more problematic regions and faulty cells in the forward part of the ECAL, this will to a larger extent affect the efficiency of mostly forward final states, such as di-jets. Thus, a luminosity-weighted efficiency-map in the $\eta - \phi$ plane will be used to weight each MC event. This is generated by applying the luminosity-weighted method on all reconstructed electron objects of the skimmed MC samples. Electron objects in the crack or with $|\eta| > 2.47$ are rejected. The resulting map is shown in Figure 5.2.

There are more problematic regions and faulty Front-End Boards in the negative η-range of the ATLAS detector than the positive. As a result, more negative-η electrons are associated with

Run	Map	$\mathcal{L}$ (pb^{-1})
$<$ 165589	161730	3.1376
$<$ 166142	165589	3.67505
$<$ 166497	166142	4.5809
$<$ 166658	166497	0
166658	166658	1.95275
$<$ 167521	166497	4.0847
$\geq$ 167844	167521	21.5634

Table 5.6: List of OQ maps, corresponding runs and luminosities. Only runs $\geq$ 158045 (from period D) are included, see Section 5.7.5. This includes all luminosity blocks which meet the requirements listed in Section 5.6.1.

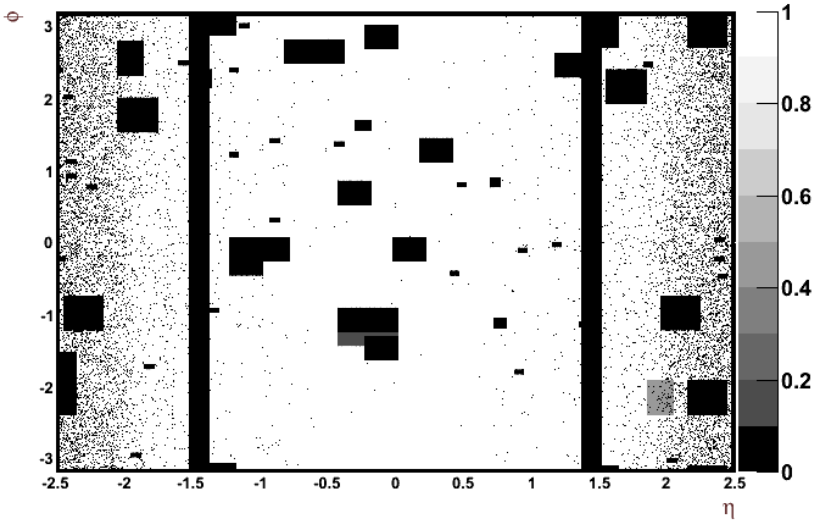


Figure 5.2: OQ efficiency map for MC. The thick, black lines (crack regions) and the black dots represent empty bins, i.e. no MC statistics. The black rectangles are problematic regions as of Map 161730, and are thus empty bins as well. To the lower right and in the middle at $\phi = -1.3$ there are two gray areas. These represent regions which failed during the 2010 data-taking.

a bad cluster than positive-η electrons. This can be seen as an asymmetry in the η-distribution, see Figure 5.3.

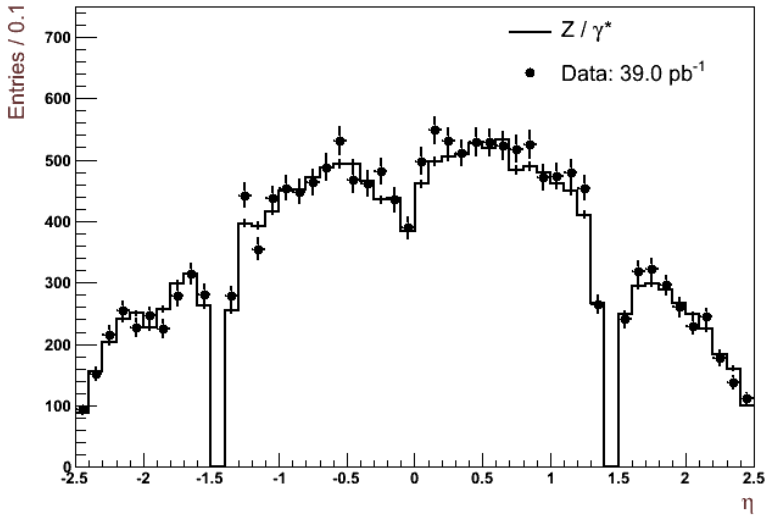


Figure 5.3: The η-distribution for data and Z/γ^* MC of the two electron candidates after applying all cuts and corrections in this section (5.7) and normalizing MC to data, see Section 5.11.1. The skewness is due to the greater occurrence of detector problems in the negative η-region.

5.7.4 Corrections to MC

Trigger simulation

A small trigger inefficiency was observed for the 2010 data. For Release 15, trigger efficiencies normalised to offline cuts for the EF e15_medium and e20_loose triggers were measured to be above 99% in the trigger plateau[3]. This is improved to 99.5 ± 0.5% in Release 16, and it is recommended to apply this scale factor to MC [85]. The difference between requiring trigger on the MC and just applying the scale factor is very small and well within the uncertainty of 0.5%.

Energy resolution smearing

Even after applying the energy rescaling on collision data, see section 5.7.5, it was found that MC doesn't correctly reproduce the $Z \rightarrow ee$ mass distribution. As it was found that MC well reproduces the $J/\Psi \rightarrow ee$ mass for central electrons, the discrepancy is attributed to the constant term C in the resolution,

$$\frac{\sigma}{E} = \frac{S}{\sqrt{E}} \oplus C.$$

Thus a smearing procedure, provided by the EnergyRescaler tool [86], is applied to the cluster energies in MC. By default the smearing tool introduces a constant term of 0.7%. In this analysis, the event number was used as the random seed.

Identification efficiency

Tag-and-probe methods have been used to measure the electron identification efficiencies. In general, the "tag" is a well-identified electron from a Z, whereas the "probe" is the other electron

[3]The trigger plateau is defined as 5 GeV or more above the HLT threshold.

from the same Z. How the tag and probe are defined depends on the method. The EGamma group has carried out the identification efficiency measurements separately in 8 η bins. In addition, the Z' group computed scaling factors accounting for the B-layer requirement as well. The resulting scaling factors are given in Table 5.7. In this analysis, the product of the scaling factors is applied as an event weight. Studies indicate that the reconstruction efficiency in MC can be used with a systematic uncertainty of 1.5% [85].

η range	scaling factor
$[-2.47, -2.01]$	0.990
$[-2.01, -1.52]$	0.982
$[-1.37, -0.80]$	0.980
$[-0.80, -0.00]$	0.984
$[+0.00, +0.80]$	0.984
$[+0.80, +1.37]$	0.988
$[+1.52, +2.01]$	0.998
$[+2.01, +2.47]$	1.005

Table 5.7: η-dependent event scaling factors for MC accounting for medium identification and B-layer requirements.

5.7.5 Collision data

Energy scale corrections

The EGamma group has precisely measured the electromagnetic energy scale in a study of $Z \to ee$ events with 2010 data [87]. They determined energy scale corrections for 50 η bins for central electrons and 8 for forward electrons. By means of the EnergyRescaler tool, the corrections are applied to the real data cluster energies.

Data periods and triggers

During the initial LHC data-taking periods, there was, in the case of highly saturated signals, a problem with the bunch crossing identification mechanism (BCID), a mechanism which assigns a calorimeter signal to the correct bunch crossing. It uses a finite impulse response (FIR) filter to extract the signal pulse amplitude and a peak finder (PF) algorithm to perform the peak identification in the linear regime. For saturated pulses, a dedicated algorithm can be used. The saturated pulses correspond to transverse energies above ~ 255 GeV. During periods A to C, the algorithm optimized for identification of saturated pulses was disabled. For signals with E_T above 800 GeV per trigger tower in the EM barrel, this is a problem. Studies have shown that this configuration will most likely incorrectly select the bunch crossing 25 ns above the correct one. Hence, the A to C periods will not be used in the analysis. With their good-run-integrated luminosity being a mere 17 nb^{-1}, this is acceptable.

For the first periods of data, the Level 1 L1_EM14 trigger is used, whereas the event filter EF_e20_loose is used for the later data. These triggers are chosen due to their high efficiency and not being prescaled[4]. In Table 5.8 the data periods used are listed along with their corresponding run numbers, trigger used and good-run-integrated luminosity.

[4]Prescaled triggers only allow for a certain fraction of triggered events to pass. The prescaling factor may also change from run to run.

Run periods	run numbers	trigger	$\mathcal{L}$ (pb^{-1})
MC		L1_EM14	
D-E	158045-161948	L1_EM14	1.32
F-I	162347-167844	EF_e20_loose	37.67
D-I	158045-167844		38.99

Table 5.8: Data periods used for the di-electron channel, listed with corresponding run numbers, trigger used and good-run-integrated luminosity.

5.7.6 Cut flows

As a benchmark, a comparison to the Z' cut flow [88] for data was made. Due to the pre-selection, it can not be compared before the electron quality cut, from which and onwards the cut flows are in perfect agreement, see Table 5.9.

Cut	events (this analysis)	events (Z' group)
Pre-selection	86131	
+ GRL	75403	
+ Vertex	75385	
+ Trigger	58906	
+ Author (two electrons)	58122	
+ η (two electrons)	54409	
+ p_T (two electrons)	34200	
+ OQ map (two electrons)	33231	
+ Medium (two electrons)	9909	9909
+ B-layer (two electrons)	9095	9095
+ Invariant mass $\geq$ 70 GeV	8803	8803

Table 5.9: Cut flow for data in the di-electron channel.

The cut flow for the c1 signal datasets is given in Table 5.10. It should be noted that the boosted Z sample is an inclusive 1-lepton sample, i.e. it does not just contain $Z(\rightarrow ee) + X$ events. This justifies what appears to be a bad acceptance.

Cut	e^+e^-	$Z + X$	$2 \times \{\tau, b, t, W\}$	$\gamma\gamma$	$e + \mu$	$e/\mu + X$	$X + X$
Pre-selection	8926	11318	5129	966	8324	8654	1656
+ Vertex	8924	11318	5129	966	8324	8649	1656
+ Author (two electrons)	8924	10905	5102	966	5386	6111	1645
+ η (two electrons)	8682	10818	5081	921	5185	5863	1638
+ p_T (two electrons)	8519	9508	4831	884	2113	3718	1579
+ OQ Map 161730 (two electrons)	7824	9154	4670	815	1924	3369	1487
+ Medium (two electrons)	7290	3356	153	40	16	1651	16
+ B-layer (two electrons)	7140	3254	140	1	7	1186	10
+ Invariant mass $\geq$ 60 GeV	7140	3192	140	1	7	1186	7
Acceptance (%)	71.4	5.74	0.37	0.02	0.07	1.98	0.023

Table 5.10: Cut flow for signal with conservation of B and L (c1) in the di-electron channel with no weights applied.

5.8 Preparing for muon analysis

In the di-muon analysis, the standard $Z' \to \mu\mu$ selection for Release 16 reconstructed data and MC, given in Table 5.11, is used. The isolation cut will however be somewhat modified, and this is explained in Section 5.8.1. What follows is a description of the cuts and corrections applied to muons in data and MC, as well as data periods and triggers used. Cut flows for data and signal (c1) are given at the end.

Cut type	variable	see Section
Primary vertex with more than two tracks	vx_n, vx_nTracks, vx_type	4.1.5
Good run list		5.6.2
Trigger		5.8.5
Combined muon	mu_isCombinedMuon	5.8.5
$\|\eta\|$ less than 2.4	mu_eta	
MS and ID hits		5.8.3
p_T larger than 25 GeV	mu_pt	
Isolation	mu_ptcone30, mu_pt	4.1.6, 5.8.1
Impact parameters	mu_d0_exPV, mu_z0_exPV	4.1.5
Invariant mass ≥ 70	mu_eta, mu_phi, mu_pt	

Table 5.11: The event selection for di-muons (standard Z' selection for Release 16 reconstructed data and MC). The full variable names are obtained by replacing "mu" with "mu_muid".

5.8.1 Isolation

The amount of QCD background in data can be evaluated by counting the number of same-charged muon pairs, see Section 5.11.2. Applying the nominal selection in Table 5.11, including isolation, on a data control sample $60 < M_{\mu\mu} < 150$ GeV yields only one same-charged muon-pair. By leaving out the isolation cut, this is increased to 49. Clearly, the isolation cut is important for suppressing QCD background, but as was shown in Section 4.1.6, both track- and calorimeter-based isolation had a catastrophic effect on the boosted Z signal. However, as can be seen from Figure 5.4, the signal di-muons only fail the cut if they are separated by less than 0.3 in ΔR. Muons will thus be required to be isolated (ptcone30/$p_T < 0.05$) unless the di-muon candidate is separated by less than 0.3 in ΔR.

5.8.2 Impact parameters

By selecting muon pairs according to the nominal selection in Table 5.11, but omitting the impact parameter cuts, the distributions for transverse (d_0) and longitudinal (z_0) impact parameters in Figure 5.5 were obtained. It can be seen that the distributions are ever so slightly asymmetric for data. The cause of this effect is unknown, but there is no obvious reason for this to affect the analysis.

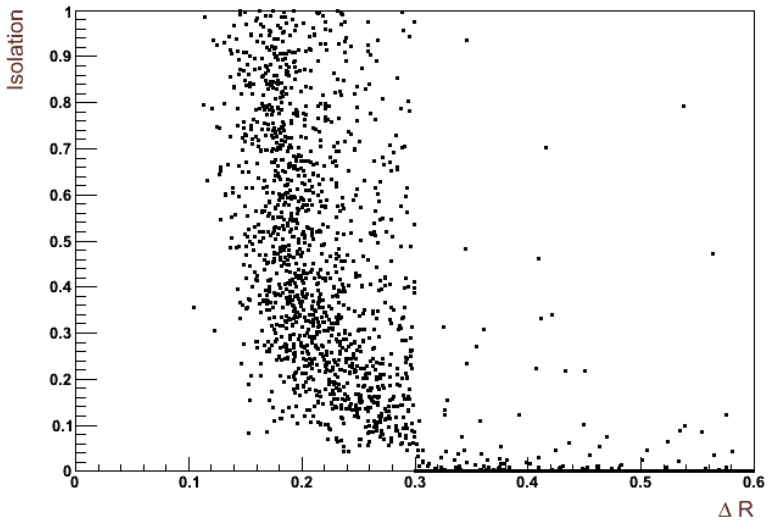
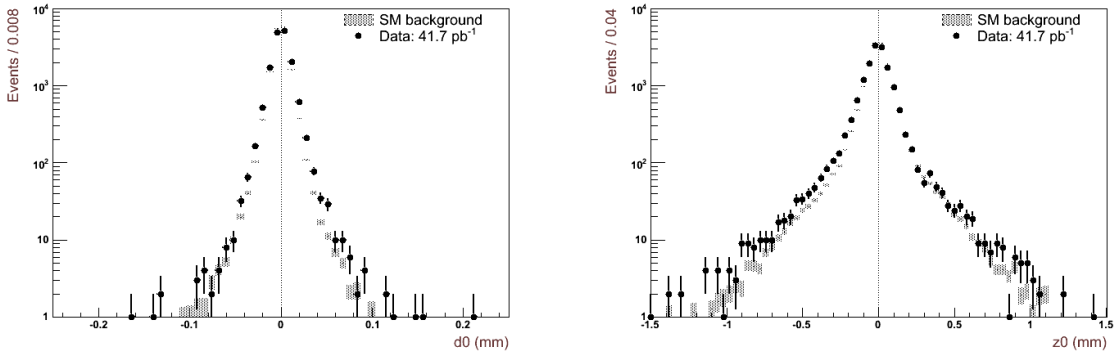


Figure 5.4: Normalized track-based isolation (ptcone30) of di-muon candidates w.r.t. separation in ΔR in the $Z + X$ (c1) signal sample.



(a) Transverse impact parameter

(b) Longitudinal impact parameter

Figure 5.5: Distributions of impact parameters of the two muon candidates after applying all cuts, apart from the impact parameter cuts, and corrections in this section (5.8) and normalizing MC to data, see Section 5.11.2.

5.8.3 Track quality

The Muon Combined Performance (MPC) group have assembled a set of recommended criteria for the associated inner detector and muon spectrometer tracks [89]. As can be seen in Figures 5.6 and 5.7 MC and data agree very well with respect to the distributions of the number of hits in the different sub-systems. The recommendations for the ID and MS tracks will thus be used throughout this analysis.

The inner detector criteria are constructed such that the condition of the silicon detectors (Pixel and SCT) are taken into account. If a passed module is dead, it is added to the hit count, but ignored in the hole[5] count. Unless the muon track passed a dead area of the innermost Pixel layer, the B-layer, a hit is required. If the track is well within the η-acceptance of the TRT, one expects at least 6 hits and the number of outliers[6] to be less than

$$0.9 \times n = 0.9 \times \left(n_{TRT}^{outliers} + n_{TRT}^{outliers} \right).$$

According to the recommendation for the associated muon spectrometer track, the muon must have at least one ϕ hit in any of the RPC, TGC or CSC layers and at least three hits in either each of the MDT barrel layers or in each of the middle and outer MDT end-cap layers and either the CSC η or MDT inner end-cap layer. The misaligned BEE, EE and BIS78 chambers are vetoed. In Table 5.12 these and the ID requirements are given more clearly.

Mostly due to the layering and alignment of the muon spectrometer, shown in Figure 5.8, the efficiency of the hits cuts is not uniform in the $\eta - \phi$ plane. The barrel consists of three concentric cylindrical layers of eight large and eight small chambers, and in Figure 2.5 one sees a similar structure of eight large and eight small chambers in the end-caps. This eight-fold structure is clearly reflected in the efficiency of the hits cuts, see Figure 5.9(a). One can also see evidence of the gap at $\eta = 0$ and the transition regions between barrel and end-caps. Here, the muon identification rate is non-existent (white).

The overall efficiency of the cuts on hits for single, isolated muons is 78.9% and 79.5% in MC and data, respectively, where the loss in acceptance is mainly due to the MS hits requirements in Table 5.12. However, more muons than expected satisfy the hits requirements, and the overall data-to-mc ratio is 1.0079±0.0011. In Figure 5.9(b) one sees that the efficiency for central muons is better than expected, but worse for forwarded muons. With the amount of data at hand, a finer binning results in bins with zero data entries. To avoid MC events being falsely weighted by zero, a coarse binning is necessary. Thus, the map in Figure 5.9(b) will henceforth be used to weight each MC event.

5.8.4 Corrections to MC

Efficiency scale factors

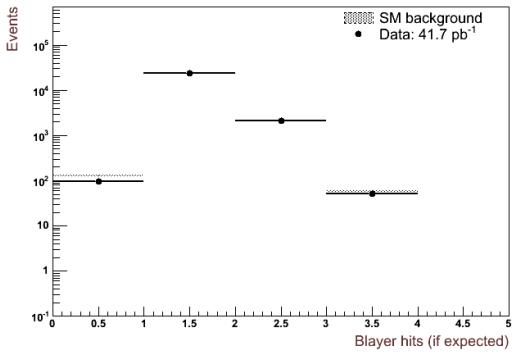
The muon efficiency is very well predicted in the ATLAS Monte Carlo. By means of the MuonEfficiencyCorrections tool provided by the MCP group, the efficiency from simulation will be scaled to the one measured from data. The systematic uncertainty on the scale factors is 0.2%.

[5] Holes on a track represent missing measurements between the first and last actual measurements. Many holes on a track is usually a strong criterion for a bad track candidate.

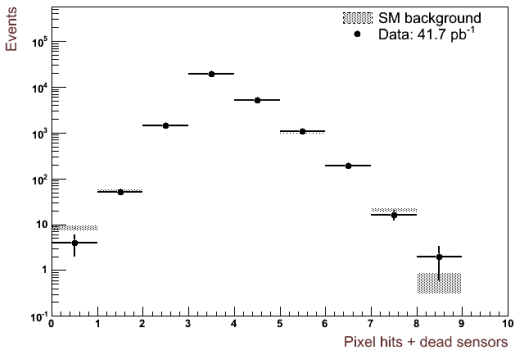
[6] Outliers are measurements which have been selected as not compatible with the track hypothesis. They can appear on misaligned detector layers, after significant energy loss, or if a second track interferes the drift tube measurements.

Cut type	variable	requirement
MDT/CSC precision layers	nMDTBIHits + nMDTBMHits + nMDTBOHits OR nMDTEMHits + nMDTEOHits + (nMDTEIHits OR nCSCEtaHits)	At least three hits in each station.
Veto BIS78, BEE and EE	nMDTBIS78Hits, nMDTBEEHits, nMDTEEHits	No hits.
Phi-hits	nRPCLayer1PhiHits + nRPCLayer2PhiHits + nRPCLayer3PhiHits + nTGCLayer1PhiHits + nTGCLayer2PhiHits + nTGCLayer3PhiHits + nTGCLayer4PhiHits + nCSCPhiHits	At least one phi-hit in either station.
BLayer	expectBLayerHit, nBLHits	At least one hit if expected.
Pixel hits	nPixHits + nPixelDeadSensors	At least two hits.
SCT hits	nSCTHits + nSCTDeadSensors	At least six hits.
Pixel + SCT holes	nPixHoles + nSCTHoles	Not more than one hole.
TRT $\|\eta\| < 1.9$	nTRTHits + nTRTOutliers	At least six hits and outliers. Outlier fraction less than $0.9 \times n$.
TRT $\|\eta\| \geq 1.9$	nTRTHits + nTRTOutliers	If number of hits and outliers is greater than five, require outlier fraction less than $0.9 \times n$.

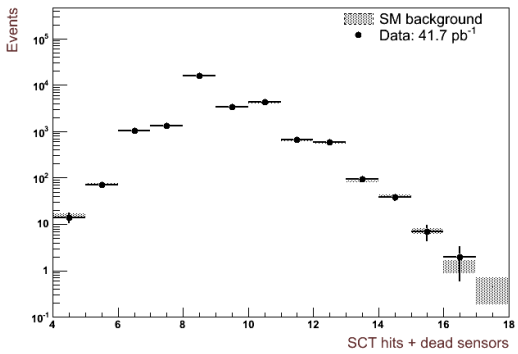
Table 5.12: The muon hits requirements recommended by the MCP group for Release 16 reconstructed data and MC.



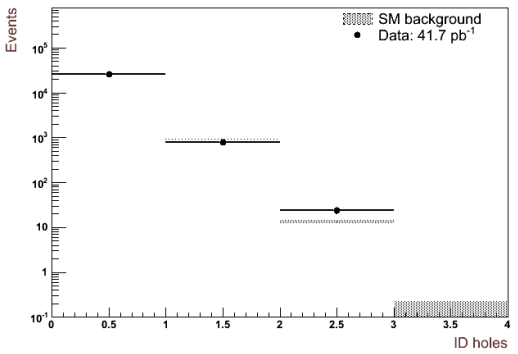
(a) Blayer hits if expected



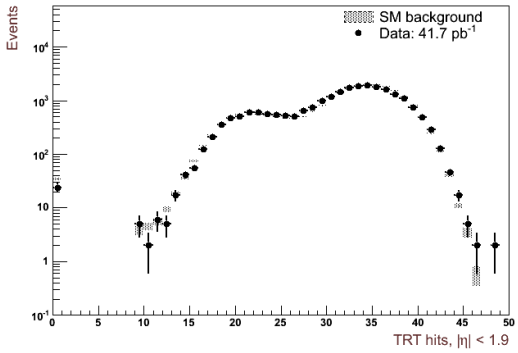
(b) Pixel hits



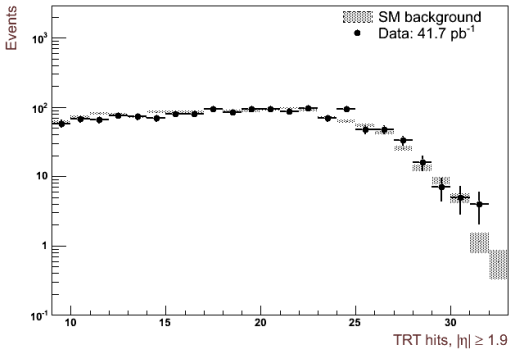
(c) SCT hits



(d) ID holes

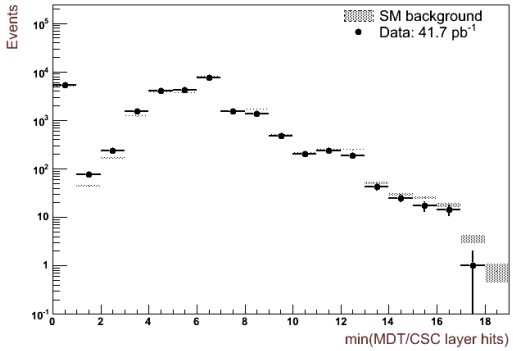


(e) TRT, $|\eta| < 1.9$

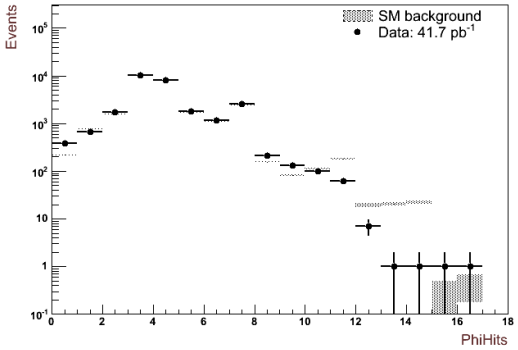


(f) TRT, $|\eta| \geq 1.9$

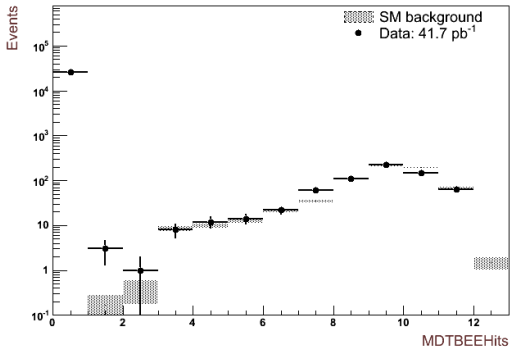
Figure 5.6: Distributions of hits in the inner detector of the two muon candidates after applying all cuts and corrections in this section (5.8) and normalizing MC to data, see Section 5.11.2.



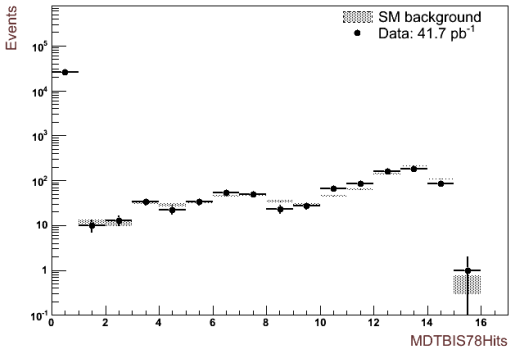
(a) Minimum hits in bending plane layers



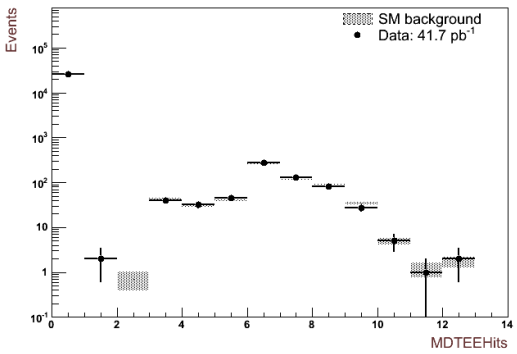
(b) Hits in phi-layers



(c) BEE hits

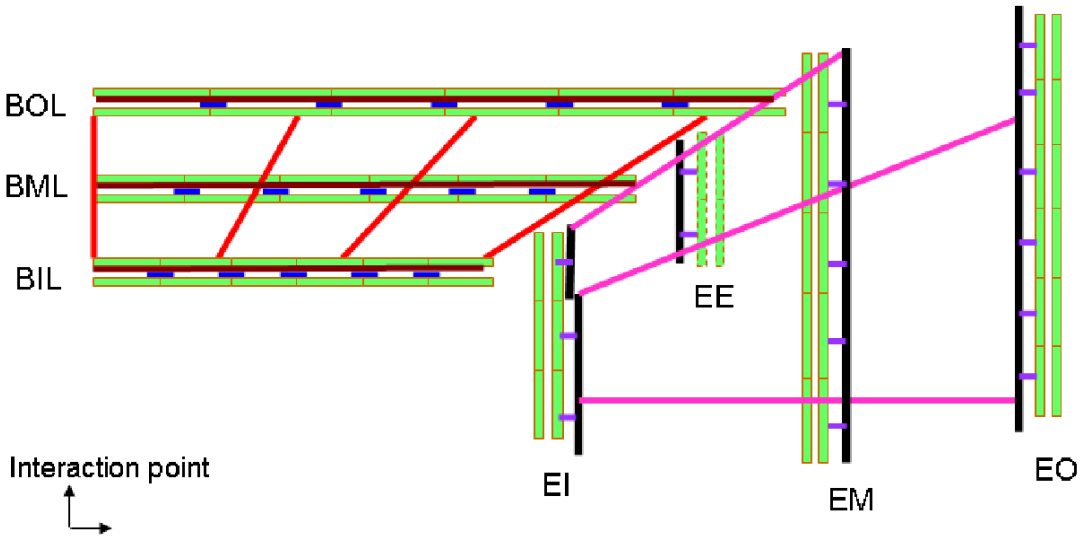


(d) BIS78 hits

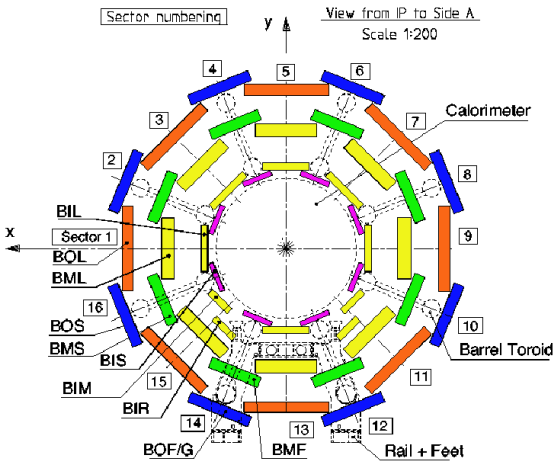


(e) EE hits

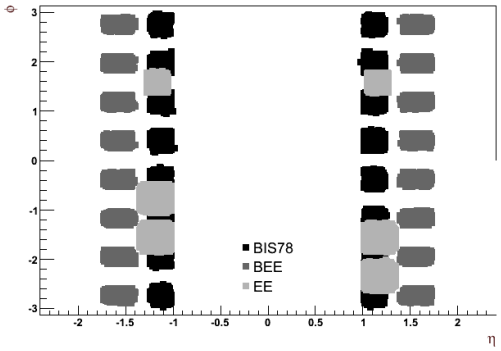
Figure 5.7: Distributions of hits in the muon spectrometer of the two muon candidates after applying all cuts and corrections in this section (5.8) and normalizing MC to data, see Section 5.11.2. In (a) the maximum of the minimum (max-of-min) hits in the barrel or end-cap layers is shown. Effectively this means that all muons with a max-of-min value ≥ 3 pass the MDT/CSC precision layers cut. The number of phi-layer hits is shown in (b), while the numbers of hits in the veto chambers are shown in the three last plots.



(a)

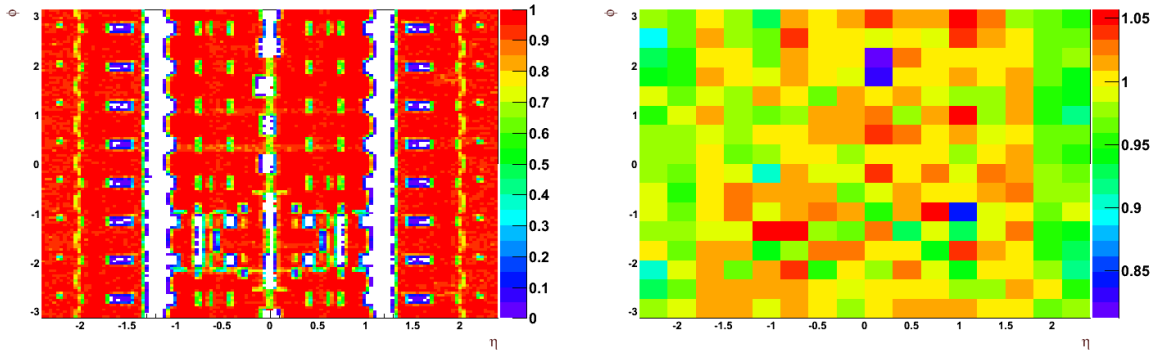


(b)



(c)

Figure 5.8: Side view of the ATLAS muon spectrometer (a) and cross-section of the barrel muon system perpendicular to the beam axis (b), showing three concentric cylindrical layers of eight large and eight small chambers. Figures (a) and (b) are from [35], but the bottom text of (a) has been cut away. The $\eta - \phi$ positions of the BEE, EE and BIS78 veto chambers are shown in the last plot (c).



(a) (b)

Figure 5.9: Efficiency of the hits cuts in MC (a), where one can see the feet of ATLAS reflected in the low-efficiency structure in the lower middle of the plot. In (b) the ratio between efficiency of the hits cuts in data and MC is shown.

Momentum resolution smearing

An evaluation of the muon momentum resolution showed that the resolution in data was worse than expected from MC [90]. The MuonMomentumCorrections tool applies scale shifting and resolution smearing on MC muons, allowing the muon momentum resolution in simulation to model closely the one in data. The tool separately smears the ID, extrapolated MS and combined track momenta and shifts the combined scale. In this analysis, the event number and muon index will be used to generate the random seed in ROOT,

```
gRand->SetSeed(680049 + EventNumber + MuonIndex*100).
```

5.8.5 Collision data

STACO vs. MuidCB

The STACO algorithm combines an inner detector track with a muon spectrometer track using a statistical method, while the MuidCB algorithm uses a global refit of the two tracks, see Section 3.4.2. In the autumn reprocessed data, which will be studied in this analysis, the STACO muons suffer from a severe extrapolation bug. The energy-loss in the calorimeters was drastically underestimated, causing a catastrophic loss of low-p_T muons. Instead of an actual average energy loss in the calorimeter of 3 GeV, the extrapolator expected 1 GeV. The effect of the bug is expected to be suppressed with increasing p_T. In this analysis the MuidCB collection will be used for both data and MC.

Data periods and triggers

Before period G5 there was an alignment problem with the combined muon event filter trigger algorithm. This caused a degraded efficiency for the trigger chains used. Thus, from period E4 one takes logical OR of the Muon and MuGirl[7] triggers. In Table 5.13 the data periods are listed along with their corresponding run numbers, trigger(s) used and good-run-integrated luminosity.

[7]MuGirl is an algorithm seeded by ID tracks. It extrapolates them into the MS and searches for matching hits.

Run periods	run numbers	trigger	$\mathcal{L}$ (pb^{-1})
MC		L1_MU10	
A-E3	152166-160879	L1_MU10	0.8
E4-G4	160899-165818	EF_mu10 OR EF_mu10_MG	6.0
G5-I1	165821-167576	EF_mu13 OR EF_mu13_MG	15.8
I1-I2	167607-167844	EF_mu13_tight OR EF_mu13_MG_tight	19.1
A-I	152166-167844		41.7

Table 5.13: Data periods used for the di-muon channel, listed with corresponding run numbers, trigger(s) used and good-run-integrated luminosity.

5.8.6 Cut flows

As a benchmark, a comparison to the Z' cut flow [91] for data was made. Due to the pre-selection, the cut flows can not be compared before the p_T cut, from which and onwards the cut flows are in perfect agreement, see Table 5.14. Note that the original isolation cut has been applied.

Cut	events (this analysis)	events (Z' group)
Pre-selection	32655	
+ GRL	30582	
+ Primary vertex	30222	
+ Trigger	29142	
+ Combined (two muons)	16927	
+ η (two muons)	16186	
+ p_T (two muons)	15955	15955
+ ID hits (two muons)	15469	15469
+ MS hits (two muons)	9038	9038
+ d0 (two muons)	8848	8848
+ z0 (two muons)	8815	8815
+ Isolation (two muons)	7967	7967
+ Invariant mass ≥ 70 (≤ 400) GeV	7743	7743

Table 5.14: Cut flow for data in the di-muon channel.

The cut flow with the modified isolation cut (see Section 5.8.1) for the c1 signal datasets is given in Table 5.15. As was stated in Section 5.7.6, the apparently bad acceptance for the boosted Z sample is justified by the fact that is an inclusive 1-lepton sample, i.e. it does not just contain $Z(\rightarrow \mu\mu) + X$ events. The isolation cut for the $Z + X$ sample is 98% efficient if only di-muon events with $\Delta R < 1$ are considered.

5.9 Preparing for electron + muon analysis

Within this analysis the lepton candidates are selected according to Tables 5.5 and 5.11. The previously mentioned identification efficiencies and energy corrections and smearing procedures will be applied to both lepton species. For triggering, the combination of the electron and muon triggers will be used. This is assumed to be very close to fully efficient.

Due to the BCID problem affecting electrons in the first data, see Section 5.7.5, only periods D-I will be used. The electron + muon analysis requires that the ECAL and both tracking systems are operating normally. Hence, only luminosity blocks declared good for both electrons,

Cut	$\mu^+\mu^-$	$Z+X$	$2\times\{\tau,b,t,W\}$	$e/\mu+X$	$X+X$
Pre-selection	7939	11318	5129	8654	1656
+ Primary vertex	7903	11294	5106	8617	1654
+ Trigger	7575	7276	4477	5295	1437
+ Combined (two muons)	7558	5745	3273	1828	1014
+ η (two muons)	7525	5726	3262	1811	1012
+ p_T (two muons)	7448	4926	2366	1606	643
+ ID hits (two muons)	7258	4079	1238	1103	301
+ MS hits (two muons)	4328	2588	770	607	188
+ d0 (two muons)	4327	2468	562	591	137
+ z0 (two muons)	4324	2462	551	589	135
+ Invariant mass $\geq$ 60 GeV	4324	2288	399	589	61
+ Isolation (if $\Delta R > 0.3$)	4302	2067	57	564	4
Acceptance (%)	43.0	3.72	0.15	0.94	0.013

Table 5.15: Cut flow for signal (c1) in the di-muon channel with no weights applied.

see Section 5.6.1, and muons, see Section 5.6.2, will be considered. In Table 5.16 the data periods used are listed along with their corresponding run numbers, triggers used and good-run-integrated luminosity.

Run periods	run numbers	trigger	$\mathcal{L}$ (pb^{-1})
MC		L1_MU10	
D-E3	158045-160879	L1_MU10 OR L1_EM14	0.7
E4-E7	160899-161948	EF_mu10 OR EF_mu10_MG OR L1_EM14	0.5
F-G4	162347-165818	EF_mu10 OR EF_mu10_MG OR EF_e20_loose	4.8
G5-I1	165821-167576	EF_mu13 OR EF_mu13_MG OR EF_e20_loose	13.7
I1-I2	167607-167844	EF_mu13_tight OR EF_mu13_MG_tight OR EF_e20_loose	18.0
D-I	152166-167844		37.7

Table 5.16: Data periods used for the electron + muon channel, listed with corresponding run numbers, trigger(s) used and good-run-integrated luminosity.

The cut flow for the c1 signal datasets is given in Table 5.17. As the electron triggers are more efficient than the muon triggers, the L1Calo (up to Run 159224) and Egamma (from Run 160387 and onward) streams are used for data.

5.10 Maximum likelihood estimation

In the following sections, the method of maximum likelihood will be used for three purposes. First, MC background will be normalized to data using a binned likelihood fit where the sum-of-squared-weights defines the uncertainty in each bin. Second, in order to extrapolate the background expectation to higher values of invariant mass and di-lepton p_T, fits will be performed on either data or MC. When the number of data points is small, an unbinned fit is preferred, but it is rather time-consuming for a large number of data points, as is the case for MC. However, due to the limited amount of data, a binned fit to MC may yield a more reliable result. The content of each bin is here assumed to be Poisson distributed. If the fit function does not vary much over a bin, or in other words, a sufficiently fine binning is chosen, the binned maximum

Cut	$\mu^+\mu^-$	$Z+X$	$2\times\{\tau,b,t,W\}$	$e+\mu$	$e/\mu+X$	$X+X$	Data
Pre-selection	7939	11318	5129	8324	8654	1656	86131
+ GRL	-	-	-	-	-	-	72933
+ Primary vertex	7903	11294	5106	8293	8617	1654	72837
+ Trigger	7660	11294	5106	8293	8564	1654	60125
+ Combined/author (electron + muon)	5353	7966	4898	8282	5551	1587	10585
+ η (electron + muon)	5240	7932	4891	8151	5452	1583	10341
+ p_T (electron + muon)	2077	6857	4615	7987	4535	1452	3873
+ ID and MS hits (one muon)	1943	5427	3076	6021	3132	914	2943
+ d0 and z0 (one muon)	1943	5172	2708	6011	3105	801	2635
+ OQ map (one electron)	1834	5048	2694	5763	2956	796	2467
+ Medium and B-layer (one electron)	13	660	586	5489	1787	109	321
+ Isolation (one muon)	13	156	158	5473	1756	13	137
+ Invariant mass $\geq$ 70 GeV	9	102	142	5473	1734	3	76
Acceptance (%)	0.09	0.18	0.37	54.7	2.89	0.001	-

Table 5.17: Cut flow for signal (c1) and data in the electron + muon channel with no weights applied to MC.

likelihood performs as well as the unbinned. For the binned likelihood fits a 1 GeV binning will thus be used. The benefit of binning is that it allows for a goodness-of-fit test. As will be described in Section 5.14.1, the *profile likelihood* approach will at last be used to calculate significances and exclusion limits.

There are two general methods of parameter estimation; least-squares estimation (LSE) and maximum likelihood estimation (MLE). Although LSE is in many ways useful, it has no basis for hypothesis testing or constructing confidence intervals. The MLE, on the other hand, has several attractive properties:

- Sufficiency. Complete information about the parameter of interest θ is contained in its MLE estimator.
- Consistency. The true parameter value θ_0 is recovered asymptotically, i.e. for sufficiently large data samples.
- Efficiency. The lowest possible variance of parameter estimates is achieved asymptotically as well.
- Asymptotic normality. The distribution of the MLE tends asymptotically to the Gaussian distribution with mean θ_0.
- Parametrization invariance. The obtained MLE solution is independent of the parametrization used.

Originally developed by R. A. Fisher in the 1920s, the method of maximum likelihood estimates θ_0 by finding a value of θ that maximizes $\mathcal{L}(\theta|x)$. The probability function $p(\theta)$ of the estimator $\hat{\theta}$ then represents the population that is "most likely" to have generated the observed data.

The RooFit toolkit [92] will be used for all the purposes mentioned above. This allows the user to model the expected distributions of events in a physics analysis, perform likelihood fits, produce plots and carry out pseudo-experiments. Being originally developed for the BaBar collaboration, an experiment dedicated to the study of B-mesons, the RooFit software is primarily designed as a particle physics data analysis tool. The brilliance of RooFit is that models are built in a truly object-oriented fashion. There are classes to express datasets, functions, p.d.f.s and even simple variables. This might sound overly complicated, but it makes RooFit an incredibly powerful tool.

The actual fitting is done with the ROOT implementation of MINUIT [93], the standard

tool for minimization[8] and uncertainty analysis in High Energy Physics. For all purposes, the MINUIT processor MINOS will be used to calculate parameter uncertainties. MINOS takes both parameter correlations and non-linearities into account, yielding a very reliable uncertainty estimation.

5.11 Background estimation

All backgrounds except from multi-jets and prompt photons are taken from MC. As seen in Section 4.2, the statistics of the multi-jet and photon + jets MC are quite bad. In addition, the reduction factors are very large and may be unreliable. In order to estimate the contribution from multi-jets and prompt photon production (collectively called "QCD" from now on), a data-driven template method will be followed. A data sample dominated by such events is obtained by replacing the requirement of two "tight" leptons by two "loose" leptons. The remaining cuts are applied as normal.

After obtaining the background templates, they are normalized to data by using the *extended maximum likelihood* technique [94]. In contrast to the "classical" maximum likelihood technique where the p.d.f. is assumed to be normalized, this absorbs the normalization into the list of parameters. In the following, only the normalization factors f fulfilling

$$N_{data} = f_{QCD}N_{QCD} + f_{MC}N_{MC}, \tag{5.1}$$

where N is the number of events, are allowed to vary in the fit. By normalizing the total invariant mass distribution to data, the dependency on the integrated luminosity and other mass-independent rate-like systematic uncertainties are removed. These include uncertainties on the trigger, identification and muon hits efficiencies, as well as on mass-independent K-factors. However, the shape-like uncertainties in the energy resolution and scale corrections, dependent on both η and p_T, are still present, but they will be neglected in this analysis.

5.11.1 Di-electron background

QCD events are for the most part expected to yield no "tight" leptons. Herein, "tight" electrons refer to electrons which pass medium identification and the Pixel B-layer cut, whereas "loose" electrons have to pass loose, but fail medium, identification. To avoid contamination of W + jets and $t\bar{t}$ events, which may very well contain a tight electron as well as two loose ones, events with tight electrons are vetoed in what will be referred to as the loose-no-tight selection. Thus, the loose-no-tight selection is fully orthogonal to the nominal selection (see Table 5.5). Note that for electrons the tight selection equals the nominal selection, i.e. the selection used for the signal search.

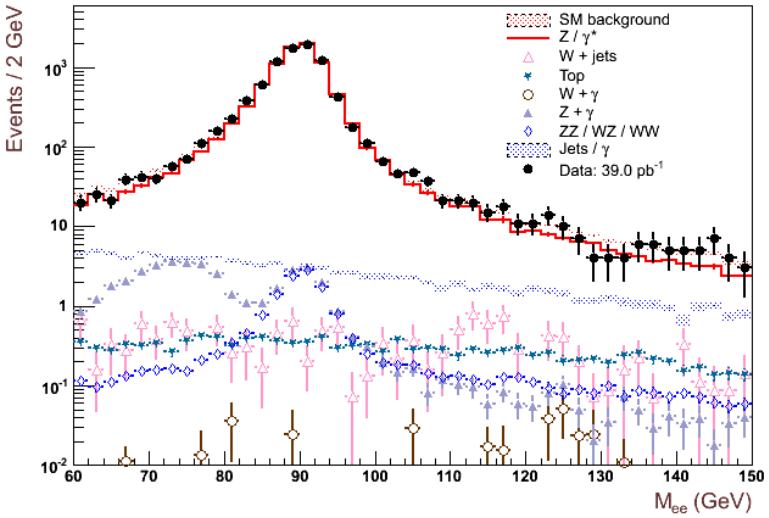
Two templates are used; (1) the sum of all MC background samples using the tight selection (sum of MC components shown in Figure 5.10(a)) and (2) a data sample derived using the loose-no-tight selection (Data in Figure 5.10(b)). These are then normalized to a data sample derived using the tight selection. The normalization factors extracted will be used to scale all other distributions. To minimize contamination from a possible signal at high masses, only events satisfying $60 < M_{ee} < 150$ GeV were considered in the fit. The results of the fit are given in Table 5.18, and good agreement between data and the estimated background can be seen in Figure 5.10(a).

The expected contamination from other backgrounds in template (2) are, before normalization, 54 Drell-Yan (0.7%), 22 W (0.3%) and 1 $t\bar{t}$ events out of a total of 7490 events, see Figure

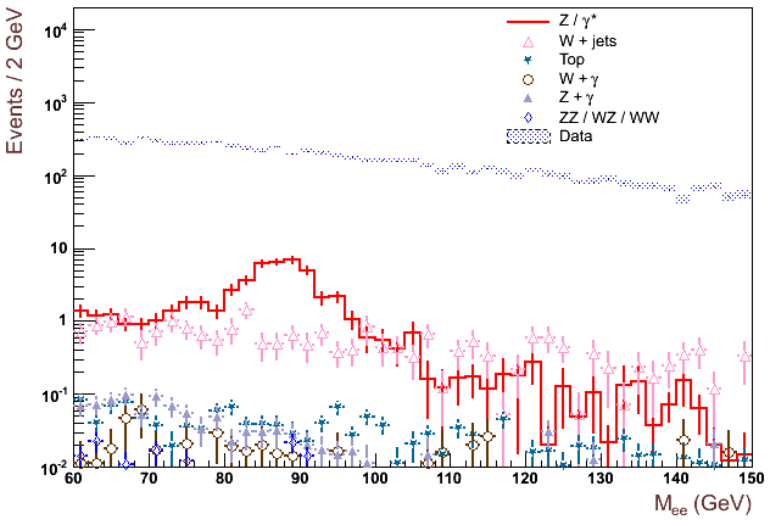
[8]The minimization of the negative log-likelihood $-2\ln\mathcal{L}$ is analogous to the maximization of the likelihood. In practice it is often more convenient to work with the log-likelihood.

Sample	events	stat. uncertainty	norm. factor
QCD (multi-jet / γ)	116	$+39/-38$	0.0155
Total MC	8770	$+100/-100$	1.31
Data	8884		

Table 5.18: The event yield estimated from data of each di-electron sample.



(a) Invariant mass distribution in data and estimated background in the di-electron channel. The SM background is the sum of the normalized MC background components and the normalized QCD sample derived from data using the loose-no-tight selection.



(b) Invariant mass distribution in data and MC after the loose-no-tight selection in the di-electron channel. The data sample describes the shape of the estimated QCD background.

Figure 5.10:

5.10(b). This is assumed to have little effect on the background estimate in the high-mass and high-p_T tails. By vetoing events with tight electrons the event yield of template (1) decreases by 0.4%, but the contamination of the MC backgrounds is reduced by 12%. Another method for suppressing W and $t\bar{t}$ events is to require the missing E_T (MET) to be below e.g. 20 GeV. However, this relies on the detector (e.g. the calorimeters) being properly calibrated. From Figure 5.11 one can see that data and estimated background agree pretty well, but there are some discrepancies, and thus the MET cut is not used herein.

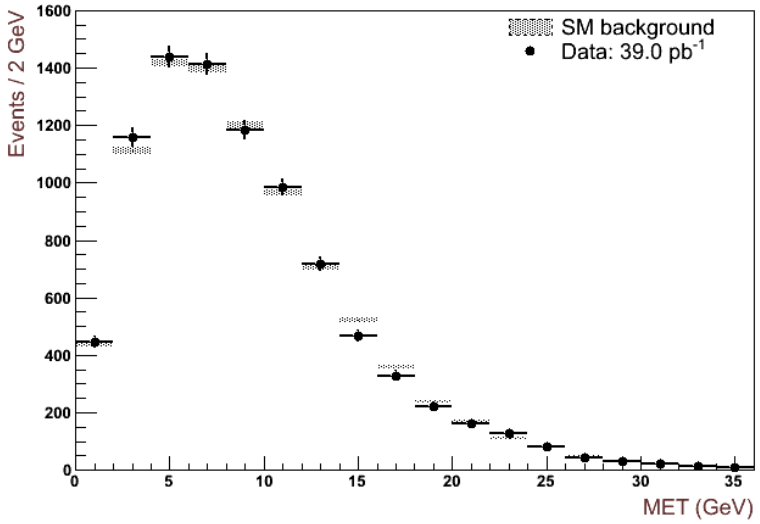


Figure 5.11: Missing E_T in data and sum of estimated backgrounds after normalization and applying the tight selection in the di-electron channel.

QCD invariant mass fit

The QCD template contains very few high-mass events and none above 1 TeV. It is thus necessary to perform a fit to give a proper estimation of the QCD background at high masses. Clearly, the efficiency of the p_T cut is lower for di-electrons of low than high invariant mass, and performing the fit in the invariant mass-region above 60 GeV would probably underestimate the QCD background at high mass. In the previous chapter, the significances for high-mass searches were estimated in the invariant mass-region above 120 GeV. The fit will therefore be performed in this mass-range, under the assumption that it is sufficiently on the plateau for the p_T cut. The loose-no-tight selection was applied to the signal samples as well as an inclusive PBH sample ($M_P = 2$ TeV), i.e. a sample containing for the most part very high-mass di-jets. It was found that the possible signal contamination in the invariant mass-region ≥ 70 GeV for $M_P = 2$ TeV is in the order of 0.1% and hence a negligible effect on the background estimation.

Given the appearance of the data, an appropriate fit function is the sum of two exponentials,

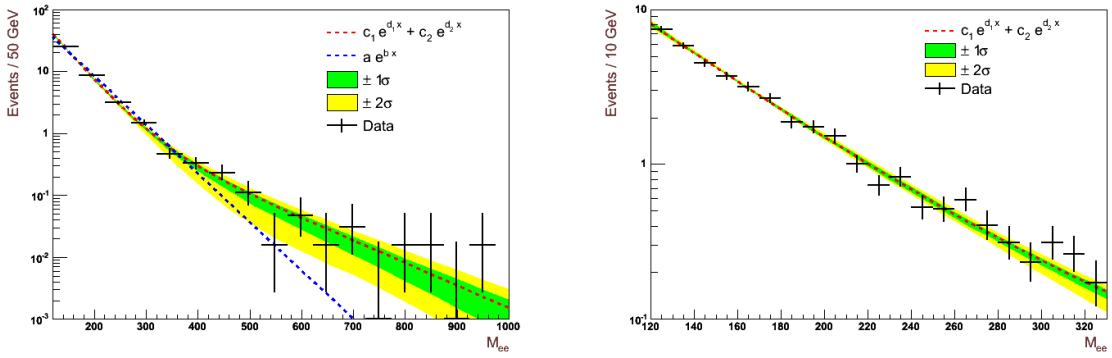
$$f(x) = c_1 e^{d_1 x} + c_2 e^{d_2 x}, \tag{5.2}$$

where the expected number of events in the mass-range (x_1, x_2) is given by

$$N = \int_{x_1}^{x_2} f(x) dx. \tag{5.3}$$

RooFit was used to perform an unbinned maximum likelihood fit, with the intent to squeeze all of the available information out of the data. In Figure 5.12(a) the fit function and binned data

are shown for invariant masses up to 1 TeV. As one can see, a single exponential function underestimates the QCD background in the invariant mass-region above 400 GeV. The uncertainty bands were calculated using a sampling method, giving a more robust uncertainty estimate. A number of curves were calculated by varying the parameter values. These were drawn from a multi-variate Gaussian p.d.f. which is constructed from the covariance matrix of the fit result. The uncertainty for each value of M_{ee} is found by calculating a central interval that captures 68% (green band) and 95% (yellow band) of the variations.



(a) (b)

Figure 5.12: The invariant mass distribution from the QCD sample fitted by a sum of exponentials (Eq. 5.2). The histogram and fit function have been normalized according to the QCD normalization factor given in Table 5.18.

By binning the data in bins of 10 GeV, the fit can be tested with Pearson's χ^2 in the invariant mass-range $(120, 330)$ GeV. The event yield of bins above this range is less than the commonly defined minimum adequate bin count for χ^2 tests, namely 10. For normally-distributed data points the χ^2/n_{dof} is typically around 1. This means that each observed bin count, on average, have a 1σ deviation from the expected. The fit function Eq. 5.2 describes the data rather well, which is verified in terms of a $\chi^2/n_{dof} = 0.54$ in the test invariant mass-range, see Figure 5.12(b).

Summary of uncertainties

The invariant mass-range $(60, 150)$ GeV was used to normalize MC to data, and statistical uncertainties were therein considered. Hence, only the invariant mass-region above 150 GeV will be used in the high-mass di-lepton search later on. The statistical uncertainty on each bin is then uncorrelated with the systematic uncertainty on the normalization factor. The shape-uncertainty on the QCD fit is somewhat correlated with the uncertainty on the normalization factor since the fit and normalization ranges overlap. They will, however, be treated as uncorrelated. The uncertainties in the mass-dependent K-factors for the Drell-Yan background given in Table 5.2 are likely to be highly correlated, and thus a part of this uncertainty is absorbed in the normalization. In the following, the K-factors will be treated as fully correlated and the uncertainties are thus given by

$$\Delta_r^+[\%]^{\text{corrected}} = \Delta_r^+[\%] \oplus \overline{\Delta_r^+}[\%]^{\text{norm}} \quad , \quad \Delta_r^-[\%]^{\text{corrected}} = \Delta_r^-[\%] \oplus \overline{\Delta_r^-}[\%]^{\text{norm}}. \tag{5.4}$$

Here, $\overline{\Delta_r^\pm}[\%]^{\text{norm}}$ is the weighted average upper and lower uncertainty in the normalization region, that is, 1.7% and -1.6%, respectively.

5.11.2 Di-muon background

For muons, the light flavour fake background can be estimated by studying the di-muon invariant mass distribution in the case where both muons carry the same charge. One expects the charge to be uncorrelated in such events. On the other hand, real muons from heavy flavour decay are more likely to be opposite-sign (OS), and special MC samples will be used to account for this background. Di-muons are herein referred to as OS if the muons are opposite-sign, and SS if they are same-sign.

In each event, the two highest-p_T muons will be selected as the di-muon candidate. Reconstructed events containing more than two high-p_T muons are in any case rare. After applying the nominal muon selection (see Table 5.11), except for the isolation and invariant mass cuts, there is only one event in the 2010 data which contains three reconstructed muons, one of which is perfectly isolated, see Table 5.19. The MET of the tri-muon event is 55.9 GeV, which makes it a good candidate for both W and top production. The semi-isolated muon may be due to heavy flavour-decay, such as bottom or charm, while the least isolated muon is most likely associated with a jet.

p_T (GeV)	charge	p_T-normalized isolation	ϕ	η
76.2	1	0.288	-2.45	-0.12
68.1	1	0	-0.593	0.0977
30.1	-1	2.23	-0.721	-0.592

Table 5.19: Properties of the muons in the tri-muon event.

Only one SS di-muon event, with an invariant mass of 74.3 GeV, was found in data in the region $60 < M_{\mu\mu} < 150$ GeV. As a cross check, the SS (require same charge) and OS (require opposite charge) selections were applied on the di-jet, charm and bottom MC samples as well. It was found that no events passed $M_{\mu\mu} > 60$. Thus, the jet background for the di-muon channel was found to be negligible. In the di-electron channel, however, the QCD background was found to be the next most important background. The large difference between the di-electron and di-muon channels is mostly due to the isolation cut used for muons. To retain the highest possible signal acceptance, opposite charge is thus not required, leaving the nominal selection unchanged. Being the only background template, the normalization factor for the sum of MC components is trivial. The statistical uncertainty was however extracted by fitting to data events in the region $60 < M_{\mu\mu} < 150$ GeV. The results are given in Table 5.20, and Figure 5.13 shows good agreement between data and MC background after normalization.

Sample	events	stat. uncertainty	norm. factor
Total MC	7830	$+89/-88$	1.258
Data	7830		

Table 5.20: The event yield of each di-muon sample after normalization.

Summary of uncertainties

There are no uncertainties on the QCD background, as it is found to be negligible. Otherwise, see Section 5.11.1.

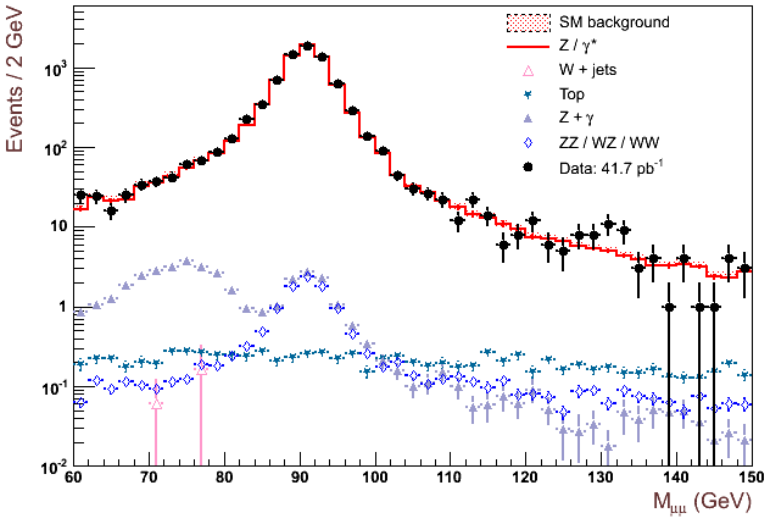


Figure 5.13: Invariant mass distribution in data and MC background (QCD is negligible) in the di-muon channel.

5.11.3 Electron + muon background

In the case of opposite-flavour lepton-pairs, "tight" electrons are defined as electrons which pass the medium identification and Pixel B-layer cuts, while "tight" muons are defined as isolated muons. The tight selection is thus identical to the nominal selection described in Section 5.9. In the loose-no-tight selection, electrons have to pass loose, but fail medium, identification, and muons must be non-isolated. Events with electrons of higher quality or isolated muons are rejected.

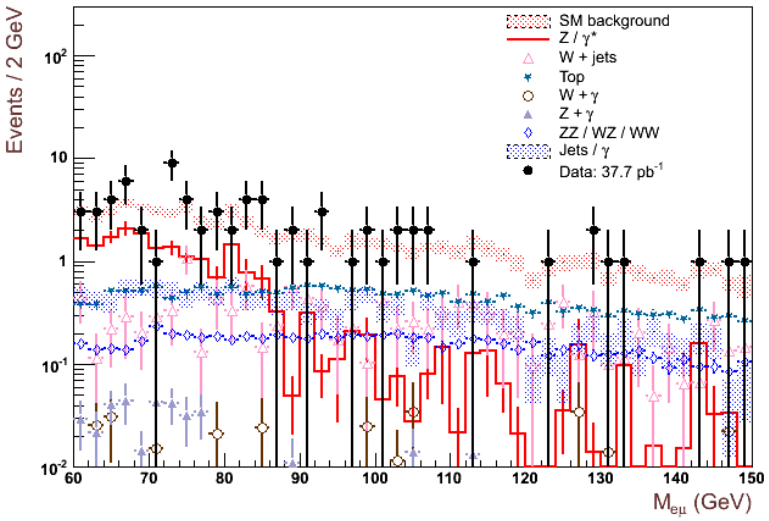
Just as for di-electrons, two templates are used; (1) the sum of all MC background samples using the tight selection (sum of MC components shown in Figure 5.14(a)) and (2) a data sample derived using the loose-no-tight selection (Data in Figure 5.14(b)). These are then normalized to a data sample derived using the tight selection in the mass-range $(60, 150)$ GeV. The results are given in Table 5.21. As can be seen in Figure 5.14(a), there are very few electron + muon events in the normalization-region; only 73 data events. Hence, the statistical uncertainties are large, but the normalization factors found are consistent with those found for the same-flavour di-lepton channels. Expected contamination from other backgrounds in template (2) are, before normalization, 7 W (1%), 2 Drell-Yan and 1 $t\bar{t}$ events out of a total of 673 events, see Figure 5.14(b).

Sample	events	stat. uncertainty	norm. factor
QCD (multi-jet / γ)	14	$+21/-14$	0.021
Total MC	59	$+20/-21$	1.2
Data	73		

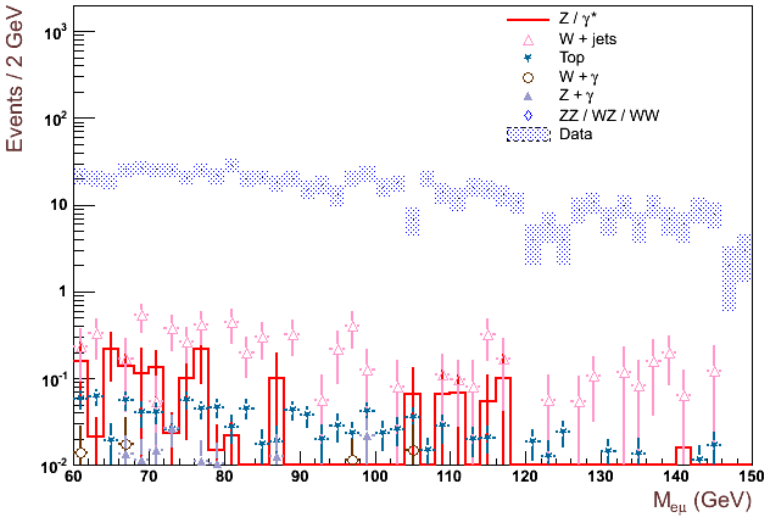
Table 5.21: The event yield of each electron + muon sample after normalization.

Invariant mass fit

The background for the electron + muon channel is tricky. In Figure 5.15(a) one can see that the available MC does not give a satisfactory description of the background for high invariant

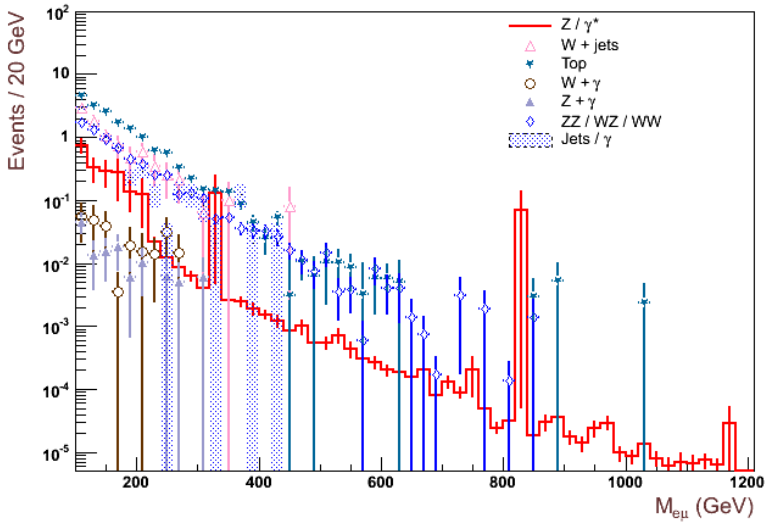


(a) Invariant mass distribution in data and estimated background in the electron + muon channel. The SM background is the sum of the normalized MC background components and the normalized QCD sample derived from data using the loose-no-tight selection.

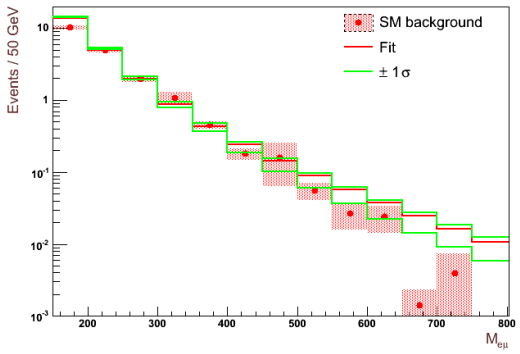


(b) Invariant mass distribution in data and MC after the loose-no-tight selection in the electron + muon channel. The data sample describes the shape of the estimated QCD background.

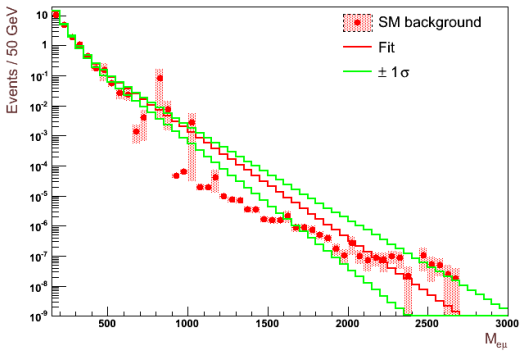
Figure 5.14:



(a) Background components in electron + muon channel



(b)



(c)

Figure 5.15: The plots in (b) and (c) show the QCD fit compared to the sum of electron + muon backgrounds.

masses. While Drell-Yan is the most important background for same-flavour di-leptons, this role is now taken by top-pair, W, di-boson and jet production, all of which have very poor or non-existent statistics for invariant masses above 400-500 GeV. Attempts were made to perform fits in different regions and with various bin widths, but none were successful. Either the fit region did not capture the second term in the double exponential (Eq. 5.2) and thus underestimated the background for very high invariant masses, or the uncertainties were too large to give a proper estimate of the parameters.

While not very accurate, it is not completely wrong to assume that the di-electron QCD invariant mass fit can be used to describe the electron + muon background for high invariant masses. Background due to heavy particle decays yields a large fraction of MET, and thus the plateau for the p_T cut is presumably at a larger value of invariant mass than in the case of QCD di-electrons. Avoiding the large fluctuation in the Drell-Yan background at $\sim 300-350$ GeV, the QCD fit is thus normalized to the sum of backgrounds in the invariant mass-range $(200, 300)$. The statistical uncertainty in this range is negligible compared to the total uncertainty in the normalization to data, that is, $+40\%/-35\%$. For invariant masses below 300 GeV, the sum of MC and QCD extracted from data will be used to describe the background expectation, while the QCD fit takes over from 300 GeV.

Although hard to judge, the QCD fit seems to nicely compensate for the lack of MC statistics for invariant masses up to 800 GeV, see Figure 5.15(b). However, if the shape of the other backgrounds matches that of Drell-Yan at very high invariant masses, the QCD fit here gives a poor estimate. As can be seen in Figure 5.15(c), the QCD fit converges to zero faster than the MC Drell-Yan. This is counter-intuitive as the initial state of a Drell-Yan process must be quark-antiquark, and the anti-quark PDFs are known to be softer than the quark PDFs. However, the QCD fit is based on a limited amount of data, and thus the extrapolation might not properly capture the shape for very high invariant masses. Another probably important effect is the degraded muon resolution for very high-p_T muons.

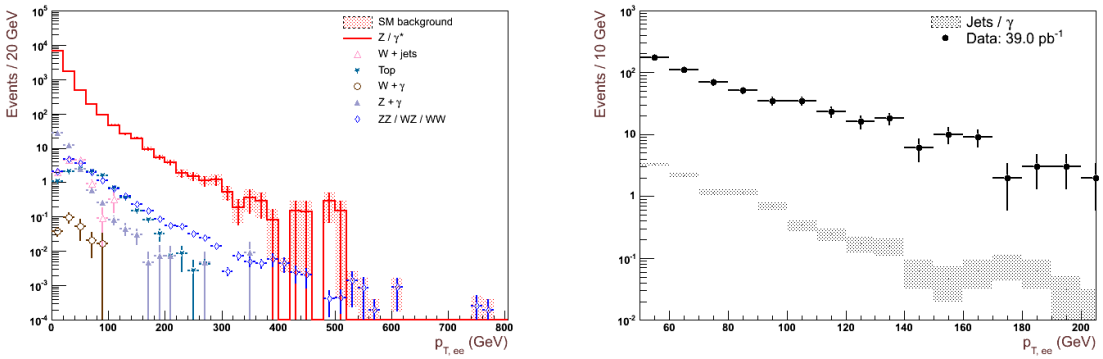
5.11.4 Di-lepton p_T fit

As can be seen in Figure 5.16, the MC statistics for highly boosted objects with mass $\in (60, 120)$ GeV is low in both the di-electron (in the figure) and di-muon channels. One can also see that the shape of the QCD template matches that of data in the di-electron channel. Thus, a fitted sum of exponentials (Eq. 5.2) will be used to describe all backgrounds in the boosted Z search. An unbinned fit was first performed on data in the range $(50, 300)$ GeV, in which a possible signal contamination is small[9]. This described the data in the fit range well, but compared to MC, it seemed to under-estimate the background at values above 300 GeV. This is illustrated in Figures 5.17(a) and 5.17(b) for the di-electron and di-muon channels, respectively. Instead, a binned fit to the sum of MC and QCD (for the di-electrons) background was performed in the range $(50, 600)$ where a reasonable amount of statistics is present. The results for the di-electron and di-muon channels are shown in Figures 5.17(c) and 5.17(d), respectively.

Summary of uncertainties

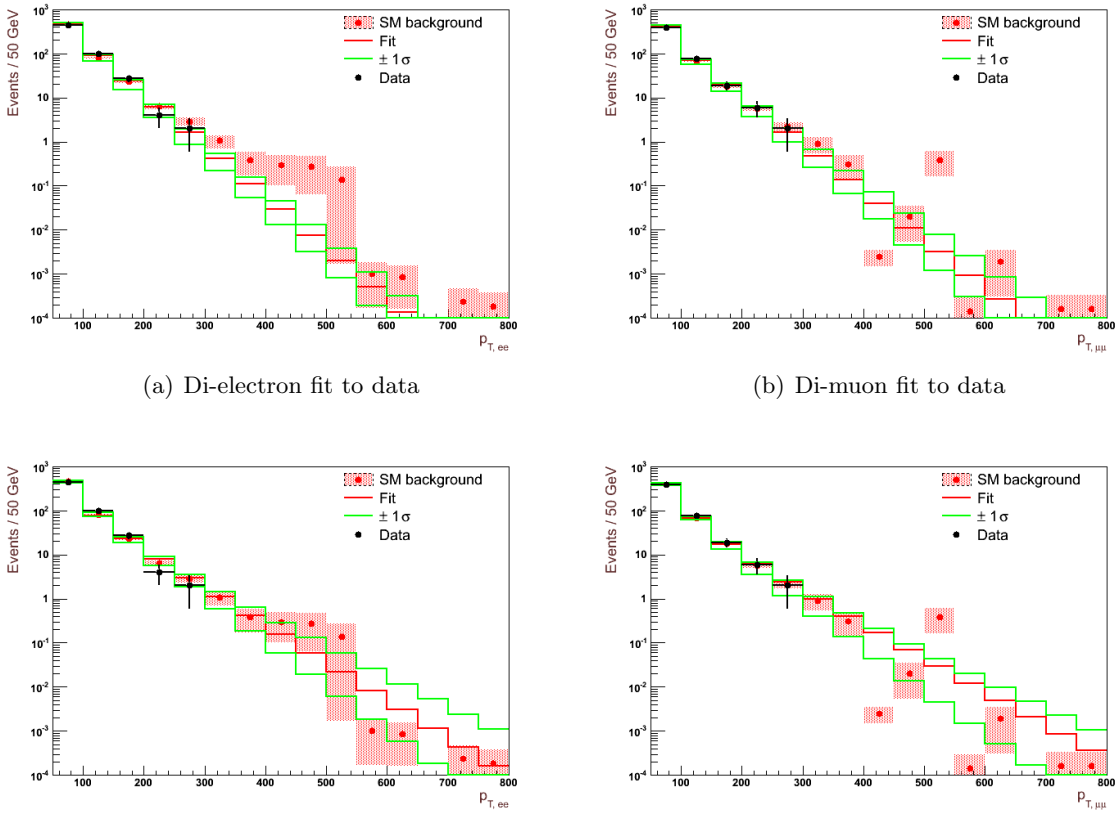
The fits were performed on a subset of the events used to estimate the normalization factors. In this mass-range, the uncertainty on the Drell-Yan K-factors is more or less constant, and it is assumed that this is absorbed and corrected for in the normalization factors found. Moreover,

[9]Due to the dependency on the PBH production threshold, it is difficult to estimate the signal contamination. In Figure 4.15 the signal di-electron p_T was shown for $M_P = 1.5$ TeV, and the contribution below 300 GeV is within the statistical uncertainty of the background. It is also unlikely that a lower Planck mass would have gone unnoticed in other searches at ATLAS.



(a) (b)

Figure 5.16: Di-electron p_T distributions in (a) MC and in (b) QCD template and data.



(a) Di-electron fit to data

(b) Di-muon fit to data

(c) Di-electron fit to MC

(d) Di-muon fit to MC

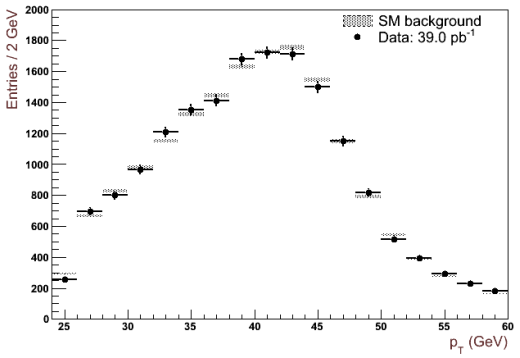
Figure 5.17: The di-electron and di-muon p_T fitted by a sum of exponentials (Eq. 5.2). Data is only shown for values up to 300 GeV. The hatched areas represent the statistical uncertainty in the SM background.

the statistical uncertainty on the normalization factors is incorporated in the shape-uncertainty from the fits, and therefore only the latter will be considered in the boosted Z analysis later on.

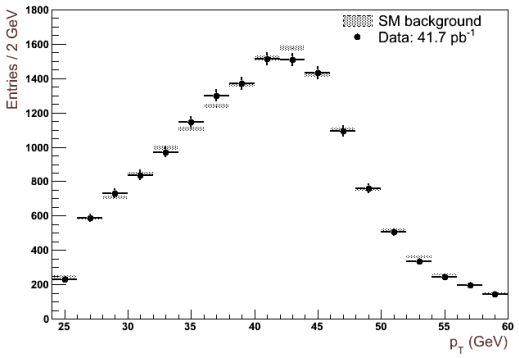
5.12 Comparison of shapes

Before proceeding it is worth taking a closer look at the invariant mass and p_T distributions in regions where data and MC *should* coincide. In the previous chapter it was shown that Drell-Yan was the most dominant background for a same-flavour di-lepton search, especially in the vicinity of the Z-peak. Drell-Yan pair-production serves as a "standard reference candle" in high-energy physics, and data and MC should therefore be in good agreement around the Z-peak.

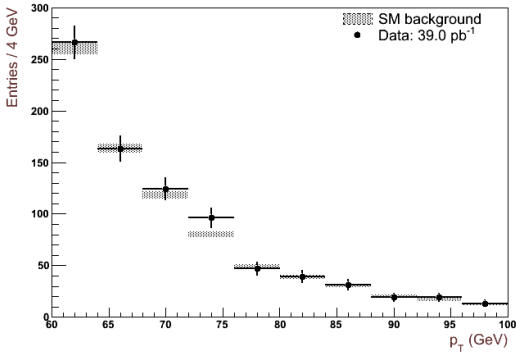
The lepton and di-lepton p_T distributions in Figures 5.18 and 5.19 show some discrepancies in data around the Z-peak (lepton $p_T \sim 45$ GeV and di-lepton $p_T \sim 5$ GeV), but the overall agreement is good. Most important for high-mass searches is the agreement at larger values, and one can see that the lepton p_T agreement is very good in the range above 75 GeV. The di-lepton p_T is not perfectly reproduced in data, but the agreement at larger values is still good.



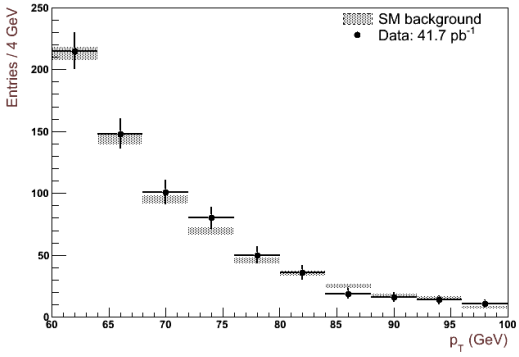
(a) Electron p_T



(b) Muon p_T

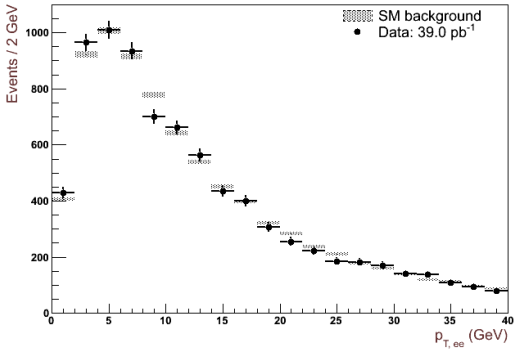


(c) Electron p_T

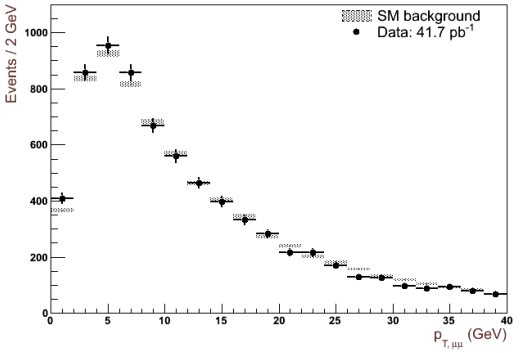


(d) Muon p_T

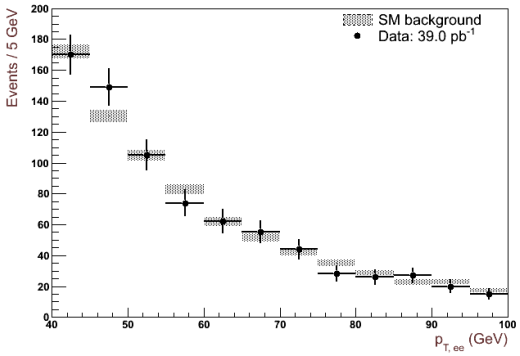
Figure 5.18: Distributions of lepton p_T after applying all cuts ($60 \leq M_{ll} \leq 150$ GeV) and corrections and normalizing MC to data.



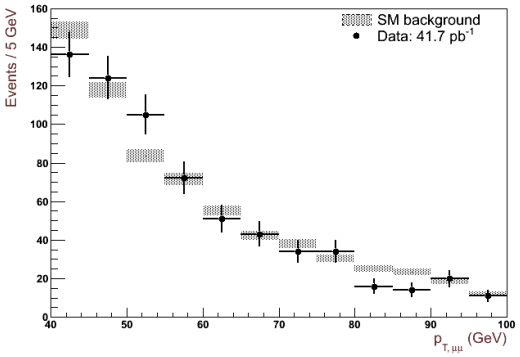
(a) Di-electron p_T



(b) Di-muon p_T



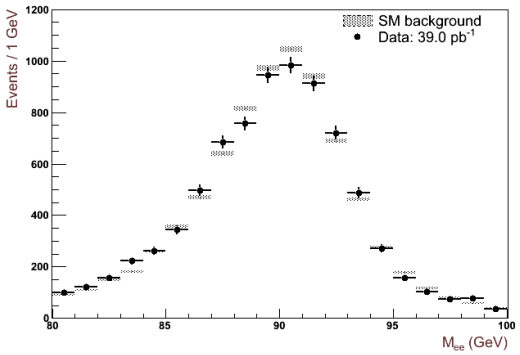
(c) Di-electron p_T



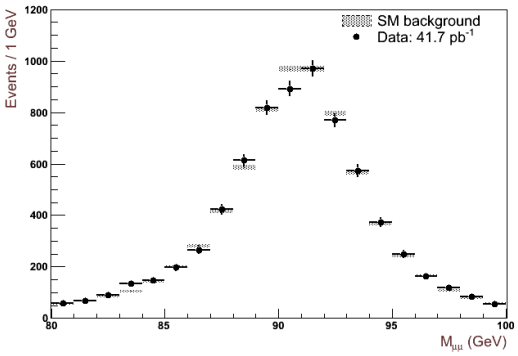
(d) Di-muon p_T

Figure 5.19: Distributions of di-lepton p_T after applying all cuts ($60 \leq M_{ll} \leq 150$ GeV) and corrections and normalizing MC to data.

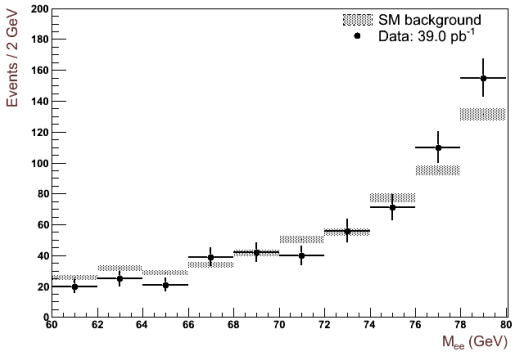
In Figure 5.20 one can see that the 2010 data does not reproduce the Z-peak very well. The shape of the data is a bit blunt in the di-electron channel and deformed in the di-muon channel compared to MC. The cause of this discrepancy is unknown, but the incomplete pile-up simulation might take some of the blame. Interactions between beam and beampipe or gas were not accounted for, but this has been improved in later MC productions. Albeit the fluctuations in data, there is a generally good agreement between data and MC in the mass-range above 110 GeV. One can see that the fluctuations in the di-electron and di-muon channels tend to cancel each other.



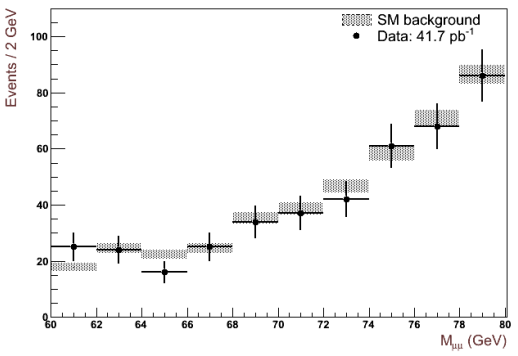
(a) Di-electron invariant mass (Z-peak)



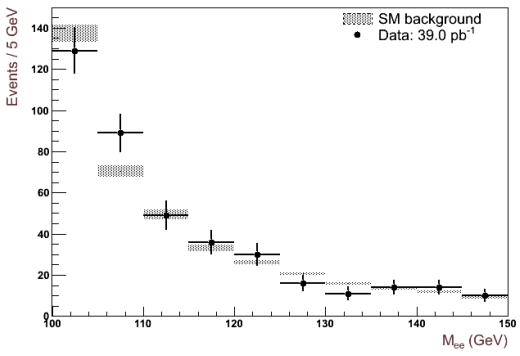
(b) Di-muon invariant mass (Z-peak)



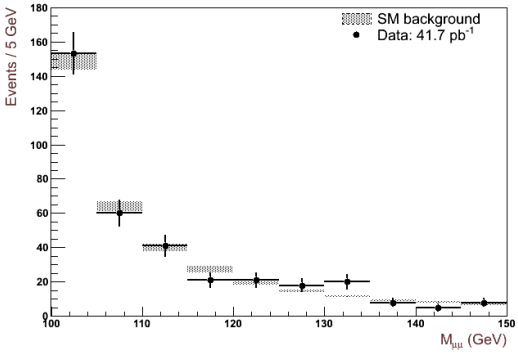
(c) Di-electron invariant mass (below Z)



(d) Di-muon invariant mass (below Z)



(e) [Di-electron invariant mass (above Z)



(f) Di-muon invariant mass (above Z)

Figure 5.20: Invariant mass distribution of same-flavour lepton pairs after applying all cuts and corrections and normalizing MC to data.

5.13 Signal templates

The $M_P = 2.0$ TeV signal shapes will be used to extrapolate to slightly lower and higher values of the Planck mass, that is, in the range $(1800, 2200)$ GeV in steps of 100 GeV. As was explained in Section 3.2.1, the branching ratios are functions of the conservation laws as well as the Planck mass. Thus, the branching ratios were calculated for each new value of the Planck mass and set of conservation laws. The two-lepton filter efficiencies showed a clear dependence on the Planck

mass, but a smaller dependence on the conservation laws. Thus, assuming that the effect of the conservation laws is negligible, the one-lepton filter efficiencies were calculated for each new value of the Planck mass with conservation of B and L (c1) only. The relative efficiencies are given in Table 5.22.

M_P (GeV)	1800	1900	2100	2200
$Z + X$	0.955	0.982	1.025	0.975
$2 \times \{\tau, b, t, W\}$	0.992	0.996	1.004	1.003
$X + X$	0.952	0.965	0.965	0.982

Table 5.22: Relative one-lepton filter efficiencies w.r.t. to $M_P = 2.0$ TeV for conservation of B and L.

Using the branching ratios and generator efficiencies, the $M_P = 2.0$ TeV signal shapes were scaled accordingly to model the signal shapes for lower and higher values of the Planck mass. However, several crude assumptions were made in the extrapolation:

- All signal-efficiencies are assumed to be constant w.r.t the Planck mass; reconstruction, trigger, event selection etc.
- The invariant mass distribution for the new Planck mass M_P^* was obtained by transforming each reconstructed invariant mass according to

$$M_{ll}^* = M_P^* + \frac{M_P^*}{M_P^0}(M_{ll} - M_P^0) \times s_{smear}, \tag{5.5}$$

  where M_{ll}^* is the new invariant mass, M_P^0 the reference Planck mass 2.0 TeV and s_{smear} a smearing/anti-smearing factor which is here set to 1. As the Z-mass is a constant, the invariant mass-shifting will not be applied to di-leptons from the boosted Z samples if $60 < M_{ll} < 120$ GeV.
- The di-lepton p_T distribution for the new Planck mass M_P^* was obtained by transforming each reconstructed di-lepton p_T according to

$$p_{T,ll}^* = \frac{M_P^*}{2} + \frac{1}{2}\frac{M_P^*}{M_P^0}(2p_{T,ll} - M_P^0) \times s_{smear}, \tag{5.6}$$

  where $p_{T,ll}^*$ is the new di-lepton p_T, and s_{smear} is here set to 1.

The signal templates are shown in Figure 5.21.

In order to investigate the sensitivity to the shape of the signal, one can apply a pessimistic 10% smearing/anti-smearing and see how this affects the final results. However, this only makes sense for a signal which shape approximately follows the normal distribution. For this reason, a special search combining only the boosted Z channels will be done. To avoid unwanted effects due to contributions from e.g. the $X + X$ channel, the templates used in the special search are purely based on the $Z + X$ sample only. The sensitivity to the shape is then evaluated by comparing the results obtained for the smeared and anti-smeared signals shapes to the nominal signal shape. This was done for $M_P^* = 1.8$ TeV and the c1 set of conservation laws. It was found that upper limits on the cross section are not affected by more than $+1.1\%/-9.6\%$. The smeared and anti-smeared signal templates, along with the nominal template for comparison, are shown in Figure 5.22.

Although not directly comparable, it was shown in Section 3.4.3 that the Release 15 Atlfast-II reconstruction efficiency is stable in the Planck mass-range $(1500, 2000)$ GeV. It is also reasonable to believe that the trigger and event selection efficiencies remain stable. The major uncertainties are those on the luminosity, namely 11%, and statistical uncertainties in each template bin.

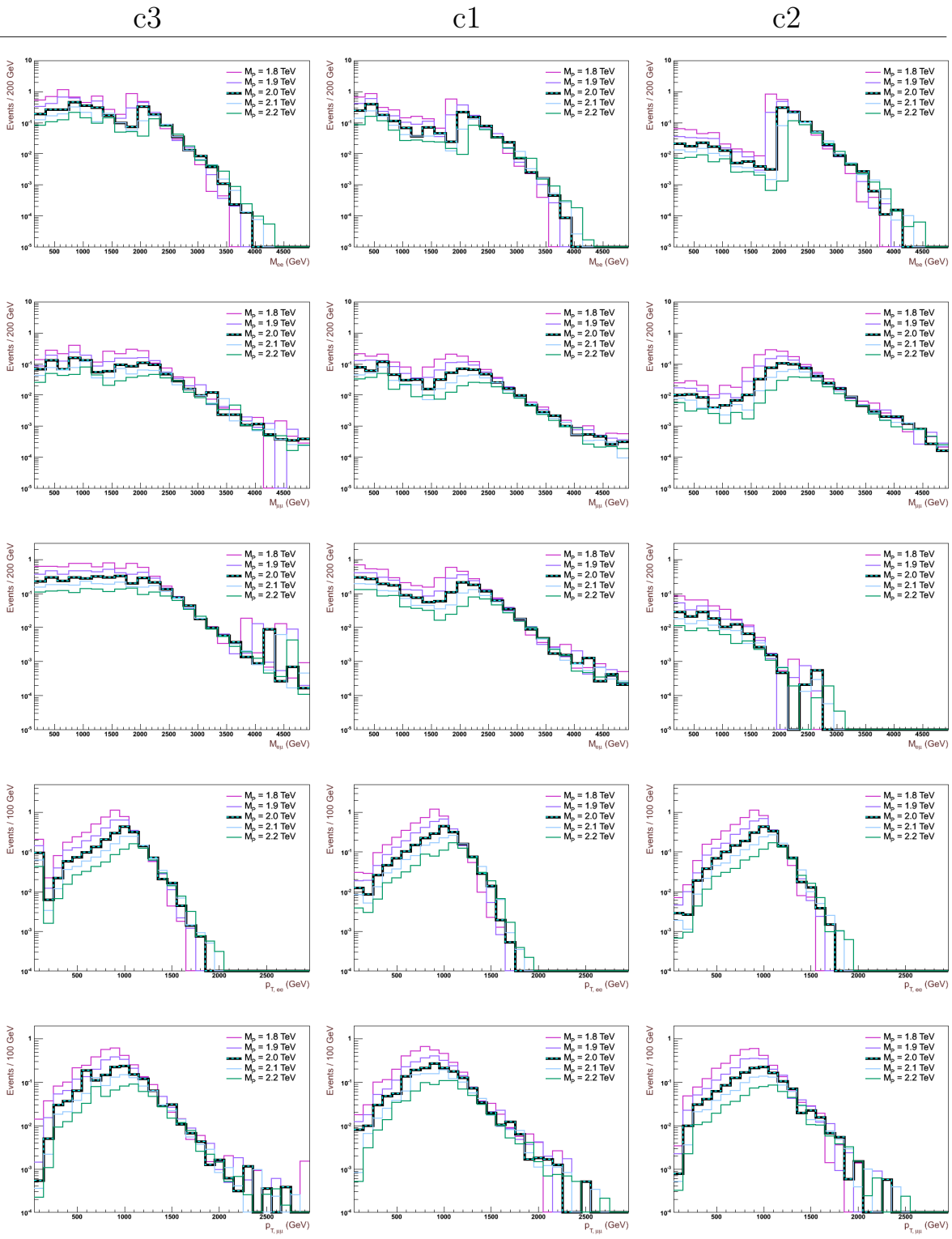
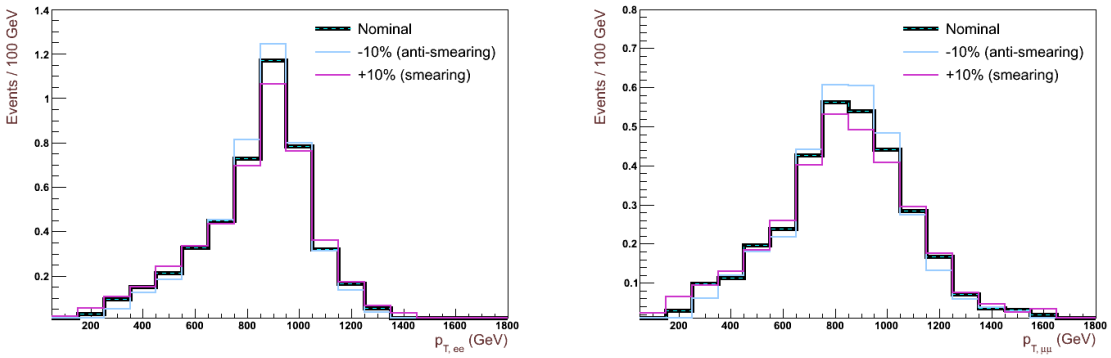


Figure 5.21: Stacked signal templates in (from top to bottom): di-electron invariant mass, di-muon invariant mass, electron + muon invariant mass, di-electron p_T and di-muon p_T spectra. The plots are ordered by the set of conservation laws (from left to right): c3 (B-L), c1 (B and L) and c2 (B, L and flavours). The reference template ($M_P = 2.0$ TeV) in each plot is marked by a thick black line with cyan dots.



(a) Di-electron p_T (b) Di-muon p_T

Figure 5.22: Di-lepton p_T smeared, nominal and anti-smeared templates for the $Z+X$ channel only. The templates are for $M_P^* = 1.8$ TeV and the c1 set of conservation laws.

Compared to this, all the other systematic uncertainties discussed in the preparation sections are, being in the order of 0.1-1%, only minor corrections. This also holds for the systematic uncertainties on the cross section, branching ratios and one-lepton filter. Thus, for the sake of simplicity, only the uncertainty on the luminosity will be included in the signal models.

5.14 Exclusion limits

As one can see from the plots in Figure 5.23, there are no signs of high-mass states in the di-lepton channels with the 2010 data. The agreement with the background expectation is verified in terms of a $CL_b = 0.40$, which was calculated using the profile likelihood method described below. It is thus appropriate to set upper limits on the signal cross section at 95% confidence level, and for this the power-constrained procedure will be followed.

5.14.1 The profile likelihood method

In the presence of systematic uncertainties, there are several independent parameters in addition to the parameters of interest (POIs). While the values of the other parameters are typically of no interest and are thus called *nuisance* parameters, the POIs are the quantities one wants to estimate or obtain confidence intervals for. The most common frequentist approach to incorporate the nuisance parameters is to do a series of pseudo-experiments, where from one pseudo-experiment to another, the values of the nuisance parameters are allowed to vary within their assumed distributions. This is, however, computationally very expensive.

The profile likelihood method reduces the number of parameters by writing them as a function of the POIs. The function is the nuisance parameter values which maximizes the likelihood given the value of the POIs. In other words, the nuisance parameters δ are replaced by its MLE (see Section 5.10) for each fixed value of the POIs θ and thus defining the profile likelihood $\mathcal{L}_{pl}$ as

$$\mathcal{L}_{pl}(\theta) = \mathcal{L}(\theta, \delta(\theta)) = \max_{\delta} \mathcal{L}(\theta, \delta)\,. \tag{5.7}$$

This method requires significantly less computing time, often as much as two orders of magnitude less than the pseudo-experiment approach [95].

The simplest example of systematic uncertainties that can be represented by nuisance parameters in profile likelihoods are rate-uncertainties. A highly relevant example is the uncertainty

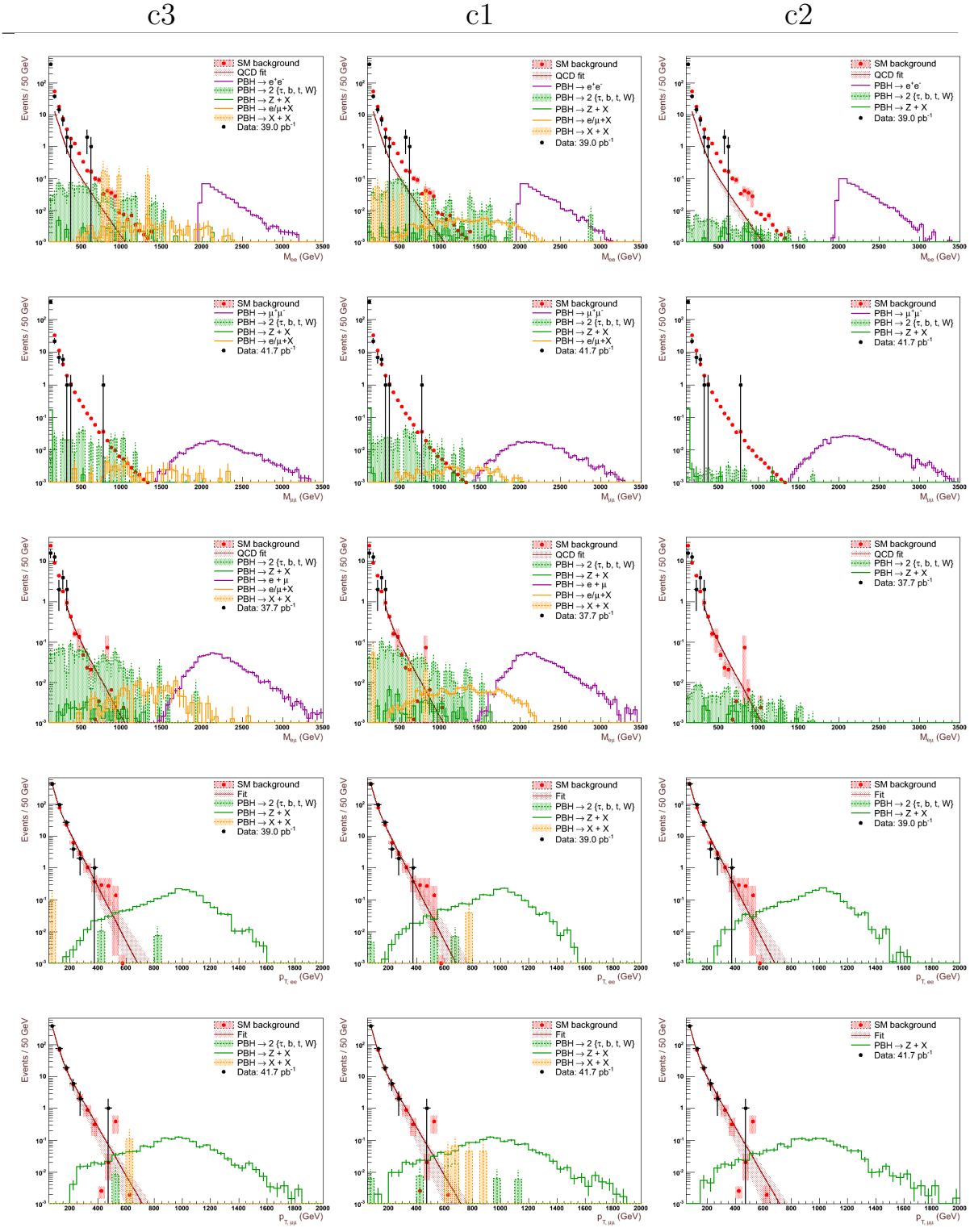


Figure 5.23: Signal and background vs. data in (from top to bottom): di-electron invariant mass, di-muon invariant mass, electron + muon invariant mass, di-electron p_T and di-muon p_T spectra. The signal samples are for $M_P = 2.0$ TeV and ordered by the set of conservation laws (from left to right): c3 (B-L), c1 (B and L) and c2 (B, L and flavours). The SM background histograms contain the sum of MC background and data-driven QCD estimate, while the fit functions are drawn as curves. See Section 5.11 for how the QCD estimates and fits were obtained.

in the luminosity. The number of expected events in a channel or bin is typically given by

$$\mu_i = \sum_{j=1}^{n} L\sigma_j\varepsilon_{ji}, \tag{5.8}$$

where L is the integrated luminosity, σ_j the cross section of source j and ε_{ji} the efficiency for source j in channel or bin i. Assuming that the measured luminosity follows a Gaussian distribution, the likelihood is then given by

$$\mathcal{L} = \prod_{i=1}^{N} \mathcal{P}(n_i|\mu_i)\mathcal{G}(L|\tilde{L}, \sigma_L), \tag{5.9}$$

where $\mathcal{P}(n_i|\mu_i)$ is the Poisson likelihood for a statistically independent bin or channel, L the measured luminosity and $\mathcal{G}$ a normalized Gaussian of mean $\tilde{L}$ and width σ_L. The negative log-likelihood is then simply given by

$$-\ln\mathcal{L} = \sum_i [-n_i \ln\mu_i + \mu_i] + \frac{(L-\tilde{L})^2}{2\sigma_L^2}, \tag{5.10}$$

such that the Gaussian term acts as a penalty. It is possible to use other functions than the Gaussian as well.

5.14.2 Power-constrained upper exclusion limits

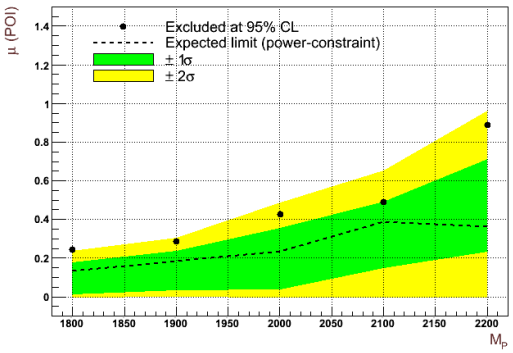
In the following, combined channel upper limits on the signal cross section at 95% confidence level will be calculated. The POI μ is defined as

$$\mu = \sigma_{PBH}^{observed}/\sigma_{PBH}^{theory},$$

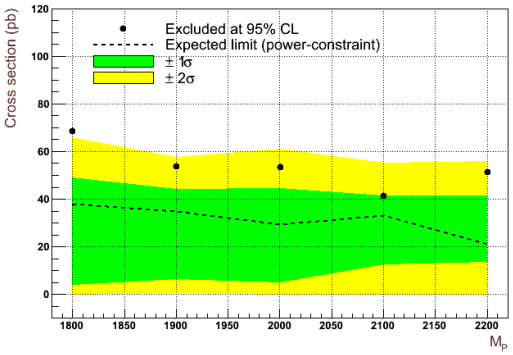
and its maximum likelihood estimator $\hat{\mu}$ is assumed to follow a Gaussian distribution. For this task, the power-constrained procedure with a threshold $M_{min} = \Phi(0) = 0.5$ will be followed[10]. This is the most recent recommendation from the ATLAS statistics forum. Here, Φ is the standard normal distribution with a mean of zero and unit standard deviation. Thus, the power constraint will be imposed whenever the observed limit is found below the median upper limit on μ under the background-only hypothesis. This is appropriate for low statistics searches, as one is prohibited from taking advantage of a fortunate downward fluctuations in the data. The distribution of the expected upper limit is found by doing a series of pseudo-experiments for $\mu = 0$. From this the 68% and 95% bands as well as the median expected limit are extracted.

The results are shown in Figure 5.24. As one can see, the limits are worse than expected, and some points even surpass the upper 95% band. This is most likely due to an upward fluctuation in the data. As a test, by removing the five most signal-like data points, the observed limit on μ for $M_P = 2.0$ TeV and the c1 set of conservation laws is brought down to 0.26. This is below the median constraint, but within the lower 68% band. The five most signal-like data events are the three di-electron events with an invariant mass above 500 GeV, the di-muon event with an invariant mass above 750 GeV and the di-muon event with a di-muon p_T above 450 GeV. From Figure 5.23 one clearly sees that these events exceeds the background expectation. However, in the case of signal, one would expect to see an excess of events in the signal-enriched regions as well and not just in the upper tails of the background. As there is no evidence that a signal is present, the excesses are ascribed to statistical fluctuations.

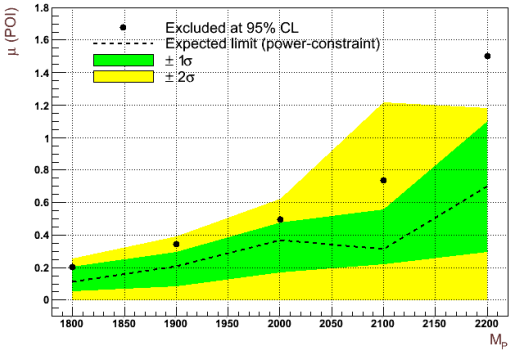
[10]The original article recommended using a threshold of $M_{min} = \Phi(-1) = 0.1587$, such that M_{min} represented a fluctuation of one standard deviation below the median [68].



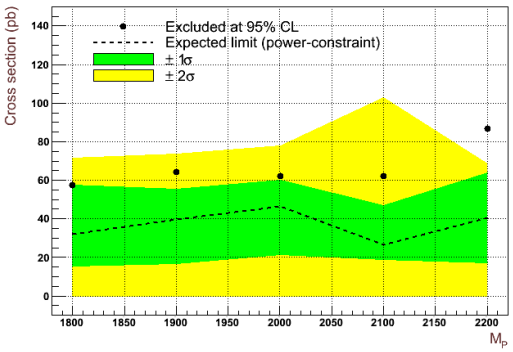
(a) 95% confidence level limit on μ (c3)



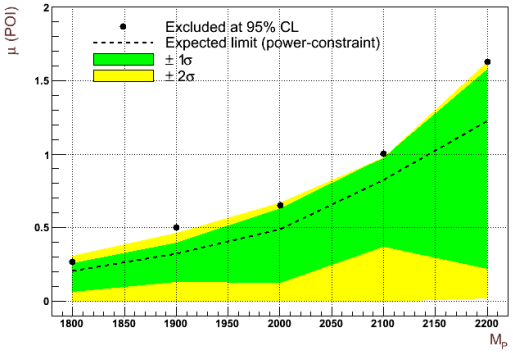
(b) 95% confidence level limit on PBH production cross section (c3)



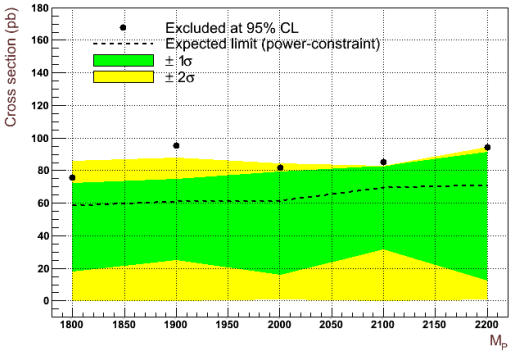
(c) 95% confidence level limit on μ (c1)



(d) 95% confidence level limit on PBH production cross section (c1)



(e) 95% confidence level limit on μ (c2)



(f) 95% confidence level limit on PBH production cross section (c2)

Figure 5.24: 95% confidence level limits on the PBH production cross section divided by the cross section predicted by theory are shown in the left plots. 95% confidence level limits on the PBH production cross section in units of pb are shown in the right plots. The plots are ordered by the set of conservation laws in the signal model (from top to bottom): c3 (B-L), c1 (B and L) and c2 (B, L and flavours).

The limits on the extrapolated signal models are only indicative. However, upper limits have been set for PBH production assuming six large extra dimensions and a Planck mass of 2 TeV. Remember that the two-particle final state scenario is general model not including effects such as fermion splitting, rotation, brane tension and recoil. Neither was the effect of inelasticity accounted for. With these simplifications, upper limits at 95% confidence level on the production cross section are; 8.2×10^1 pb for conservation of B, L and flavours; 6.2×10^1 pb for conservation of B and L; 5.3×10^1 pb for conservation of B-L only.

Conclusions and outlook

In this thesis, the possibility for discovery of Planckian black holes in the two-particle final state scenario in several di-lepton channels with the ATLAS detector at the LHC has been investigated. Such processes can be seen as an excess of events in the tails of the expected di-lepton spectra. No such excess was observed, however.

Corrections to data, as well as weighting and smearing procedures for MC, were applied accordingly to the recommendations of the combined performance groups at ATLAS. In addition, the data cut flows was found to be in perfect agreement with the ones obtained by the Z' group at ATLAS. Due to the poor MC coverage for some of the relevant background processes and spectra, fits where performed in order to extrapolate the background expectation to the signal-enriched regions.

By means of the profile likelihood method, the most dominant systematic uncertainties were incorporated. With the 2010 data taken at 7 TeV center-of-mass pp collisions, upper limits on the cross section at 95% confidence level were obtained for three Planckian black hole models. Assuming six large extra dimensions and a Planck mass of 2 TeV, the quoted limits are; 8.2×10^1 pb for conservation of B, L and flavours; 6.2×10^1 pb for conservation of B and L; 5.3×10^1 pb for conservation of B-L only.

With the rapidly increasing amount of ATLAS data, uncharted territories are constantly being unveiled, pushing the potential for discovery or limits on gravitational effects even further. The 15th of June this year, the LHC exceeded 1 fb^{-1} of delivered luminosity. These are very exiting times indeed.

Final thoughts

The event selection used is not optimized for PBH search, but adopted from the Z' group at ATLAS. Although the Z' share many similarities with PBHs decaying directly to two leptons, I should have done a more thorough study of the event selection, especially the use of isolation and opposite charge cuts. These reduce the jet background, but also signal from (semi)-leptonic decay chains which show as an excess in the intermediate mass range, i.e. below the Planck mass, but well above the Z mass. The trade-off between jet-rejection and maximizing the signal acceptance might well be an interesting study.

I would also have liked to study the importance of the number of muon hits as these cuts result in a dramatic reduction of the signal, about 40% more specifically. While I do not doubt that the recommendations of the muon combined performance group are soundly based and well suited for most analyzes, it would have been interesting to investigate how much these requirements can be loosened before the analysis becomes severely deteriorated, yielding unreliable results.

As far as I know, the available MC statistics for high di-lepton p_T events is very poor. Yet, we have seen that the boosted Z channel is highly important in di-lepton searches for two-particle final state PBHs. Moreover, boosted Z searches are not only relevant for black holes, but could also be for high-mass Higgs, technicolor and other exotic ideas such as excited leptons. If not already initiated, I believe simulation of Standard Model boosted Z production is something which should be done, and I would be happy to participate in this. There is also need of datasets for di-boson and top-pair production with high invariant masses as these serve as a highly important electron + muon background. While top-pair datasets for high invariant masses do exist, they are filtered on either two electrons or two muons, and are thus useless for electron + muon analyzes. At last, I should have simulated PBH samples for more than just one value of the Planck mass. The way I extrapolated to both higher and lower values is not very accurate.

As I was running out of time, I took a huge shortcut when jumping from background estimation to combined channel search. Instead, I should have done single channel searches in order to evaluate my understanding of each channel. Fortunately, the deviation from the expected limits could be ascribed to the small upward fluctuation in data.

Throughout this work I have acquired a taste for the field of statistics, of which my knowledge is rather limited. It is not with complete ease I have used the tools provided by the RooFit and RooStats packages without fully understanding what is going on. Most of all, I would like to develop a proper understanding of the different statistical approaches frequently used by the high energy physics community and how these are implemented into RooFit and RooStats.

Appendix A

List of acronyms

Here follows an alphabetically sorted list of acronyms frequently used throughout the thesis.

ADD	Arkani-Hamed, Dimopoulos and Dvali (model)
AOD	Analysis Object Data (ATLAS data format)
AMI	ATLAS Metadata Interface (tag)
ATLAS	A Toroidal LHC ApparatuS (detector)
BEE	Barrel End-cap Extra (muon spectrometer chamber)
BH	black hole
BIS78	Barrel Inner Small 78(muon spectrometer chamber)
CSC	Cathode Strip Chamber (muon spectrometer chamber)
D3PD	Derived 3 Physics Dataset (ROOT ntuple format)
DY	Drell-Yan (background, process, event)
ECAL	electromagnetic calorimeter
EE	End-cap Extra (muon spectrometer chamber)
EF	event filter (trigger)
GRL	Good Run List (of luminosity blocks)
HCAL	hadronic calorimeter
HLT	High-Level Trigger
ID	inner detector
L1	Level 1 (trigger)
L2	Level 2 (trigger)
LHC	Large Hadron Collider
MC	Monte Carlo (technique, simulation)
MDT	Monitored Drift Tube (muon spectrometer chamber)
MET	missing transverse energy (in event)
MLE	maximum likelihood-estimation
MCP	Muon Combined Performance (ATLAS working group)
MS	muon spectrometer
NLO	next-to-leading-order (cross section)
NNLO	next-to-next-to-leading-order (cross section)
OQ	Object Quality (map)
OSOF	oppsite-sign opposite-flavour (di-lepton)
OSSF	oppsite-sign same-flavour (di-lepton)

PBH	Planckian black hole
PDF	parton density function
POI	parameter of interest (profile likelihood fitting)
QCD	quantum chromodynamics (model, background, process, event)
QED	quantum electrodynamics (model)
QFT	quantum field theory
RPC	Resistive Plate Chamber (muon spectrometer chamber)
SCT	semiConductor Tracker (inner detector tracker)
SM	Standard Model (of particle physics)
TGC	Thin Gap Chamber (muon spectrometer chamber)
TRT	Transition Radiation Tracker (inner detector tracker)

Appendix B

Naming conventions of BlackMax samples

- SBx, BHx, LMx - The meanings of the abbreviations are: string ball, (semi-classical) black hole, low multiplicity[1] (two-body final state). x is an integer defining more precise parameters in BlackMax, e.g. 1 refers to non-rotating objects, while 2 refers to rotating objects.
- BM - BlackMax
- cx - Each integer x refers to a set of conservation laws. c1 means conservation of L as well as B, while c3 refers to BHs only conserving B - L. Finally, c2 means conservation of everything, including flavours.
- nx - x is the number of extra dimensions
- Mthx - x (in GeV) is the lower limit of the mass of a produced black hole or string ball
- MDx - x (in GeV) is the Planck scale in D dimensions
- MSx - x (in GeV) is the string scale (not relevant for this thesis)
- gSx - x is the string coupling (not relevant for this thesis)

[1] LM can also be interpreted as the "Lisa model" after Lisa Randall who, together with P. Meade, proposed the two-particle final state model [28].

Appendix C

Datasets

The tables contain run number, process name, cross section and luminosity for each dataset used in the thesis. At the top of each set of samples, the event generator and AMI tag is written.

C.1 Release 15 datasets

Run	dataset	σ (pb)	$\mathcal{L}$ (pb^{-1})
Pythia, AMI tag: e549_s765_s767_r1302_r1306			
105466	DYee_75M120	8.198×10^{2}	2.440×10^{1}
105467	DYee_120M250	8.716×10^{0}	2.294×10^{3}
105468	DYee_250M400	4.147×10^{-1}	4.811×10^{4}
105469	DYee_400M600	6.727×10^{-2}	2.972×10^{5}
105470	DYee_600M800	1.115×10^{-2}	1.794×10^{6}
105471	DYee_800M1000	2.744×10^{-3}	7.288×10^{6}
105472	DYee_1000M1250	9.178×10^{-4}	2.179×10^{7}
105473	DYee_1250M1500	2.490×10^{-4}	8.032×10^{7}
105474	DYee_1500M1750	7.655×10^{-5}	2.613×10^{8}
105475	DYee_1750M2000	2.636×10^{-5}	7.586×10^{8}
105476	DYee_M2000	1.533×10^{-5}	1.305×10^{9}
105477	DYmumu_75M120	8.199×10^{2}	2.439×10^{1}
105478	DYmumu_120M250	8.716×10^{0}	2.294×10^{3}
105479	DYmumu_250M400	4.157×10^{-1}	4.810×10^{4}
105480	DYmumu_400M600	6.724×10^{-2}	2.974×10^{5}
105481	DYmumu_600M800	1.116×10^{-2}	1.791×10^{6}
105482	DYmumu_800M1000	2.744×10^{-3}	7.270×10^{6}
105483	DYmumu_1000M1250	9.178×10^{-4}	2.178×10^{7}
105484	DYmumu_1250M1500	2.487×10^{-4}	8.041×10^{7}
105485	DYmumu_1500M1750	7.659×10^{-5}	2.610×10^{8}
105486	DYmumu_1750M2000	2.636×10^{-5}	7.584×10^{8}
105487	DYmumu_M2000	1.533×10^{-5}	1.304×10^{9}
105488	DYtautau_75M120	8.178×10^{2}	2.445×10^{1}
105489	DYtautau_120M250	8.661×10^{0}	2.309×10^{3}
105490	DYtautau_250M400	4.161×10^{-1}	4.805×10^{4}
105491	DYtautau_400M600	6.727×10^{-2}	2.972×10^{5}
105492	DYtautau_600M800	1.116×10^{-2}	1.791×10^{6}
105493	DYtautau_800M1000	2.744×10^{-3}	7.284×10^{6}
105494	DYtautau_1000M1250	9.178×10^{-4}	2.177×10^{7}
105495	DYtautau_1250M1500	2.490×10^{-4}	8.022×10^{8}
105496	DYtautau_1500M1750	7.655×10^{-5}	2.609×10^{8}
105497	DYtautau_1750M2000	2.636×10^{-5}	7.576×10^{8}
105498	DYtautau_M2000	1.532×10^{-5}	1.238×10^{9}
Pythia, AMI tag: e468_s765_s767_r1302_r1306			
106046	Zee	8.558×10^{2}	2.908×10^{3}
106047	Zmumu	8.557×10^{2}	2.920×10^{3}
106052	Ztautau	8.540×10^{2}	2.914×10^{3}
Pythia + MadGraph, AMI tag: e523_s765_s767_r1302_r1306			
108323	Zeegamma	1.002×10^{1}	4.979×10^{3}
108324	Zmumugamma	1.002×10^{1}	4.988×10^{3}
108325	Ztautaugamma	9.764×10^{0}	2.958×10^{4}
Pythia + photos, AMI tag: e552_s765_s767_r1336_r1306			
115039	diphotons50	5.885×10^{0}	1.952×10^{4}
115040	diphotons100	4.409×10^{0}	2.418×10^{5}

Table C.1: Release 15: binned and unbinned Drell-Yan, $Z + \gamma$ and γ + jet

Run	dataset	σ (pb)	$\mathcal{L}$ (pb^{-1})
Pythia, AMI tag: e468_s766_s767_r1303_r1306			
105009	J0_jetjet	9.857×10^{9}	1.420×10^{-4}
105010	J1_jetjet	6.781×10^{8}	2.058×10^{-3}
105011	J2_jetjet	4.099×10^{7}	3.410×10^{-2}
105012	J3_jetjet	2.194×10^{6}	6.370×10^{-1}
105013	J4_jetjet	8.770×10^{4}	1.593×10^{1}
105014	J5_jetjet	3.498×10^{3}	3.978×10^{3}
105015	J6_jetjet	3.362×10^{1}	4.009×10^{4}
105016	J7_jetjet	1.374×10^{-1}	8.190×10^{6}
Pythia, AMI tag: e583_s766_s767_r1303_r1306			
105017	J8_jetjet	6.279×10^{-6}	6.347×10^{8}
Alpgen + Jimmy, AMI tag: e511_s765_s767_r1302_r1306			
107680	WenuNp0	6.871×10^{3}	2.011×10^{2}
107681	WenuNp1	1.293×10^{3}	1.999×10^{2}
107682	WenuNp2	3.766×10^{2}	5.016×10^{2}
107683	WenuNp3	1.013×10^{2}	4.983×10^{2}
107684	WenuNp4	2.525×10^{1}	5.146×10^{2}
107685	WenuNp5	7.124×10^{0}	4.841×10^{2}
107690	WmunuNp0	6.871×10^{3}	2.017×10^{2}
107691	WmunuNp1	1.295×10^{3}	1.977×10^{2}
107692	WmunuNp2	3.761×10^{2}	4.995×10^{2}
107693	WmunuNp3	1.007×10^{2}	5.052×10^{2}
107694	WmunuNp4	2.599×10^{1}	4.998×10^{2}
107695	WmunuNp5	7.130×10^{0}	4.906×10^{2}
107700	WtaunuNp0	6.873×10^{3}	1.987×10^{2}
107701	WtaunuNp1	1.295×10^{3}	1.967×10^{2}
107702	WtaunuNp2	3.751×10^{2}	5.024×10^{2}
107703	WtaunuNp3	1.018×10^{2}	4.959×10^{2}
107704	WtaunuNp4	2.576×10^{1}	5.046×10^{2}
107705	WtaunuNp5	7.002×10^{0}	5.710×10^{2}
Herwig, AMI tag: e521_s765_s767_r1302_r1306			
105985	WW	2.960×10^{1}	2.173×10^{4}
105986	ZZ	4.599×10^{0}	2.558×10^{5}
105987	WZ	1.123×10^{1}	7.177×10^{4}
McAtNlo + Jimmy, AMI tag: e510_s765_s767_r1302_r1306			
105200	T1	1.441×10^{2}	1.247×10^{4}
Pythia + MadGraph, AMI tag: e521_s765_s767_r1302_r1306			
106001	Wplusenugamma	2.796×10^{1}	1.786×10^{3}
106002	Wplusmunugamma	2.794×10^{1}	1.790×10^{3}
106003	Wplustaunugamma	2.542×10^{1}	1.165×10^{4}
108288	Wminusenugamma	1.859×10^{1}	2.690×10^{3}
108289	Wminusmunugamma	1.859×10^{1}	2.690×10^{3}
108290	Wminustaunugamma	1.686×10^{1}	1.724×10^{4}
Pythia, AMI tag: e571_s765_s767_r1302_r1306			
108082	PhotonJet_Unbinned70	1.052×10^{3}	1.225×10^{3}
Pythia, AMI tag: e519_s765_s767_r1302_r1306			
108084	PhotonJet_Unbinned280	3.253×10^{0}	3.583×10^{6}

Table C.2: Release 15: di-jet, W + jet, massive di-bosons, top-pair, $W + \gamma$ and di-photon

Run	dataset	σ (pb)	$\mathcal{L}$ (pb^{-1})
Pythia + BlackMax ($M_P = 2.0$ TeV, c2), Atlfast-II			
006000	2Electrons	3.33×10^{-2}	1.50×10^{5}
006001	2Muons	3.38×10^{-2}	1.48×10^{5}
006002	1Z_X	3.666×10^{0}	2.414×10^{4}
006003	2Comb_b_t_e_mu_tau_W	8.325×10^{-1}	3.559×10^{4}
006004	2Photons	1.63×10^{-2}	1.49×10^{5}
006007	Other	1.209×10^{2}	8.065×10^{2}
Pythia + BlackMax ($M_P = 1.75$ TeV, c2), Atlfast-II			
006100	2Electrons	1.09×10^{-1}	4.59×10^{4}
006101	2Muons	1.07×10^{-1}	4.68×10^{4}
006102	1Z_X	1.241×10^{1}	7.809×10^{3}
006103	2Comb_b_t_e_mu_tau_W	2.693×10^{0}	1.111×10^{4}
006104	2Photons	5.29×10^{-2}	4.64×10^{4}
006107	Other	3.295×10^{2}	2.969×10^{2}
Pythia + BlackMax ($M_P = 1.5$ TeV, c2), Atlfast-II			
006200	2Electrons	3.67×10^{-1}	1.36×10^{4}
006201	2Muons	3.67×10^{-1}	1.36×10^{4}
006202	1Z_X	3.778×10^{1}	2.363×10^{3}
006203	2Comb_b_t_e_mu_tau_W	9.338×10^{0}	3.264×10^{3}
006204	2Photons	1.91×10^{-1}	1.29×10^{4}
006207	Other	9.464×10^{2}	1.035×10^{2}
Pythia + BlackMax ($M_P = 2.0$ TeV, c1), Atlfast-II			
007000	2Electrons	2.23×10^{-2}	2.238×10^{5}
007001	2Muons	2.21×10^{-2}	2.263×10^{5}
007002	1Z_X	3.022×10^{0}	2.805×10^{4}
007003	2Comb_b_t_e_mu_tau_W	1.141×10^{0}	2.179×10^{3}
007004	2Photons	1.15×10^{-2}	2.152×10^{5}
007005	ElectronMuon	4.45×10^{-2}	2.21×10^{5}
007007	Other	1.110×10^{2}	1.666×10^{3}
Pythia + BlackMax ($M_P = 1.75$ TeV, c1), Atlfast-II			
007100	2Electrons	7.15×10^{-2}	6.99×10^{4}
007101	2Muons	7.36×10^{-2}	6.79×10^{4}
007102	1Z_X	9.386×10^{0}	9.118×10^{3}
007103	2Comb_b_t_e_mu_tau_W	3.032×10^{1}	9.354×10^{2}
007104	2Photons	3.54×10^{-2}	6.90×10^{4}
007105	ElectronMuon	1.43×10^{-1}	6.94×10^{4}
007107	Other	3.038×10^{2}	6.125×10^{2}

Table C.3: Release 15: PBHs

Run	dataset	σ (pb)	$\mathcal{L}$ (pb^{-1})
Pythia + BlackMax ($M_P = 1.5$ TeV, c1), Atlfast-II			
007200	2Electrons	2.45×10^{-1}	2.04×10^{4}
007201	2Muons	2.45×10^{-1}	2.04×10^{4}
007202	1Z_X	3.118×10^{1}	2.766×10^{3}
007203	2Comb_b_t_e_mu_tau_W	8.449×10^{1}	3.466×10^{2}
007204	2Photons	1.28×10^{-1}	1.95×10^{4}
007205	ElectronMuon	4.88×10^{-1}	2.03×10^{4}
007207	Other	8.777×10^{2}	2.137×10^{2}
Pythia + BlackMax ($M_P = 2.0$ TeV, c3), Atlfast-II			
008000	2Electrons	2.28×10^{-2}	2.19×10^{5}
008001	2Muons	2.18×10^{-2}	2.29×10^{5}
008002	1Z_X	3.020×10^{0}	2.807×10^{4}
008003	2Comb_b_t_e_mu_tau_W	1.572×10^{1}	9.724×10^{2}
008004	2Photons	1.14×10^{-2}	2.19×10^{5}
008005	ElectronMuon	4.45×10^{-2}	2.23×10^{5}
008007	Other	1.066×10^{2}	1.430×10^{3}
Pythia + BlackMax ($M_P = 1.75$ TeV, c3), Atlfast-II			
008100	2Electrons	7.32×10^{-2}	6.83×10^{4}
008101	2Muons	7.22×10^{-2}	6.93×10^{4}
008102	1Z_X	9.420×10^{0}	9.056×10^{3}
008103	2Comb_b_t_e_mu_tau_W	4.102×10^{1}	3.771×10^{2}
008104	2Photons	3.64×10^{-2}	6.78×10^{4}
008105	ElectronMuon	1.44×10^{-1}	6.86×10^{4}
008107	Other	2.93×10^{2}	5.329×10^{2}
Pythia + BlackMax ($M_P = 1.5$ TeV, c3), Atlfast-II			
008200	2Electrons	2.51×10^{-1}	2.00×10^{4}
008201	2Muons	2.42×10^{-1}	2.07×10^{4}
008202	1Z_X	3.115×10^{1}	2.764×10^{3}
008203	2Comb_b_t_e_mu_tau_W	1.117×10^{2}	1.525×10^{2}
008204	2Photons	1.21×10^{-1}	2.04×10^{4}
008205	ElectronMuon	4.81×10^{-1}	2.05×10^{4}
008207	Other	8.505×10^{2}	1.880×10^{2}

Table C.4: Release 15: PBHs

C.2 Release 16 datasets

Run	dataset	σ (pb)	$\mathcal{L}$ (pb^{-1})
Pythia, AMI tag: e574_s933_s946_r1831_r1700			
105466	DYee_75M120	8.171×10^{2}	2.447×10^{1}
105467	DYee_120M250	8.689×10^{0}	2.071×10^{3}
105468	DYee_250M400	4.138×10^{-1}	4.348×10^{4}
105469	DYee_400M600	6.746×10^{-2}	2.816×10^{5}
105470	DYee_600M800	1.117×10^{-2}	1.791×10^{6}
105471	DYee_800M1000	2.728×10^{-3}	7.331×10^{6}
105472	DYee_1000M1250	9.165×10^{-4}	2.182×10^{7}
105473	DYee_1250M1500	2.494×10^{-4}	8.017×10^{7}
105474	DYee_1500M1750	7.688×10^{-5}	2.601×10^{8}
105475	DYee_1750M2000	2.600×10^{-5}	7.690×10^{8}
105476	DYee_M2000	1.533×10^{-5}	1.304×10^{9}
105477	DYmumu_75M120	8.171×10^{2}	2.447×10^{1}
105478	DYmumu_120M250	8.686×10^{0}	2.302×10^{3}
105479	DYmumu_250M400	4.143×10^{-1}	4.826×10^{4}
105480	DYmumu_400M600	6.726×10^{-2}	2.964×10^{5}
105481	DYmumu_600M800	1.117×10^{-2}	1.431×10^{6}
105482	DYmumu_800M1000	2.728×10^{-3}	7.329×10^{6}
105483	DYmumu_1000M1250	9.165×10^{-4}	2.182×10^{7}
105484	DYmumu_1250M1500	2.494×10^{-4}	8.016×10^{7}
105485	DYmumu_1500M1750	7.688×10^{-5}	2.610×10^{8}
105486	DYmumu_1750M2000	2.600×10^{-5}	7.689×10^{8}
105487	DYmumu_M2000	1.533×10^{-5}	1.305×10^{9}
105488	DYtautau_75M120	8.152×10^{2}	2.452×10^{1}
105489	DYtautau_120M250	8.666×10^{0}	2.307×10^{3}
105490	DYtautau_250M400	4.151×10^{-1}	4.817×10^{4}
105491	DYtautau_400M600	6.746×10^{-2}	2.964×10^{5}
105492	DYtautau_600M800	1.117×10^{-2}	1.700×10^{6}
105493	DYtautau_800M1000	2.728×10^{-3}	7.326×10^{6}
105494	DYtautau_1000M1250	9.165×10^{-4}	2.180×10^{7}
105495	DYtautau_1250M1500	2.493×10^{-4}	8.011×10^{8}
105496	DYtautau_1500M1750	7.688×10^{-5}	2.596×10^{8}
105497	DYtautau_1750M2000	2.600×10^{-5}	7.672×10^{8}
105498	DYtautau_M2000	1.532×10^{-5}	1.302×10^{9}
Pythia, AMI tag: e574_s933_s946_r1831_r1700			
106050	Zee_1Lepton	8.558×10^{2}	3.595×10^{2}
106051	Zmumu_1Lepton	8.559×10^{2}	3.594×10^{2}
106052	Ztautau	8.540×10^{2}	2.321×10^{3}
Pythia + MadGraph, AMI tag: e600_s933_s946_r1831_r1700			
108323	Zeegamma	1.002×10^{1}	4.989×10^{3}
108324	Zmumugamma	1.002×10^{1}	4.988×10^{3}
108325	Ztautaugamma	9.764×10^{0}	2.963×10^{4}

Table C.5: Release 16: binned and unbinned Drell-Yan and $Z + \gamma$

Run	dataset	σ (pb)	$\mathcal{L}$ (pb^{-1})
Pythia, AMI tag: e574_s934_s946_r1833_r1700			
105009	J0_jetjet	9.648×10^{9}	1.424×10^{-4}
105010	J1_jetjet	6.782×10^{8}	2.058×10^{-3}
105011	J2_jetjet	4.098×10^{7}	6.820×10^{-2}
105012	J3_jetjet	2.193×10^{6}	6.373×10^{-1}
105013	J4_jetjet	8.770×10^{4}	3.177×10^{1}
105014	J5_jetjet	3.350×10^{3}	5.883×10^{2}
105015	J6_jetjet	3.361×10^{1}	3.531×10^{4}
Alpgen + Jimmy, AMI tag: e600_s933_s946_r1831_r1700			
107680	WenuNp0	6.922×10^{3}	1.997×10^{2}
107681	WenuNp1	1.304×10^{3}	4.917×10^{2}
107682	WenuNp2	3.783×10^{2}	4.995×10^{2}
107683	WenuNp3	1.014×10^{2}	4.976×10^{2}
107684	WenuNp4	2.597×10^{1}	5.002×10^{2}
107685	WenuNp5	7.004×10^{0}	4.993×10^{2}
107690	WmunuNp0	6.920×10^{3}	1.994×10^{2}
107691	WmunuNp1	1.304×10^{3}	4.914×10^{2}
107692	WmunuNp2	3.778×10^{2}	4.974×10^{2}
107693	WmunuNp3	1.019×10^{2}	5.004×10^{2}
107694	WmunuNp4	2.575×10^{1}	5.047×10^{2}
107695	WmunuNp5	6.923×10^{0}	5.054×10^{2}
107700	WtaunuNp0	6.919×10^{3}	1.968×10^{2}
107701	WtaunuNp1	1.303×10^{3}	4.902×10^{2}
107702	WtaunuNp2	3.782×10^{2}	4.983×10^{2}
107703	WtaunuNp3	1.015×10^{2}	4.973×10^{2}
107704	WtaunuNp4	2.564×10^{1}	5.068×10^{2}
107705	WtaunuNp5	7.040×10^{0}	5.676×10^{2}
Herwig, AMI tag: e598_s933_s946_r1831_r1700			
105985	WW	2.959×10^{1}	2.156×10^{4}
105986	ZZ	4.596×10^{0}	2.570×10^{5}
105987	WZ	1.123×10^{1}	7.215×10^{4}
McAtNlo + Jimmy, AMI tag: e598_s933_s946_r1831_r1700			
105200	T1	1.441×10^{2}	1.245×10^{4}
Pythia + MadGraph, AMI tag: e600_s933_s946_r1831_r1700			
106001	Wplusenugamma	2.797×10^{1}	1.788×10^{3}
106002	Wplusmunugamma	2.794×10^{1}	1.790×10^{3}
106003	Wplustaunugamma	2.542×10^{1}	1.105×10^{4}
108288	Wminusenugamma	1.859×10^{1}	2.690×10^{3}
108289	Wminusmunugamma	1.859×10^{1}	2.690×10^{3}
108290	Wminustaunugamma	1.686×10^{1}	1.710×10^{4}
PythiaB, AMI tag: e574_s933_s946_r1831_r1700			
106059	ccmu15X	7.890×10^{4}	1.899×10^{1}
108405	bbmu15X	2.805×10^{4}	1.595×10^{2}

Table C.6: Release 16: di-jet, W + jet, massive di-bosons, top-pair, $W + \gamma$, charm-pair and bottom-pair

Run	dataset	σ (pb)	$\mathcal{L}$ (pb^{-1})
Pythia + BlackMax ($M_P = 2.0$ TeV, c2), AMI tag: e660_s933_r1831			
016000	2Electrons	3.36×10^{-2}	2.98×10^{5}
016001	2Muons	3.39×10^{-2}	2.95×10^{5}
016002	1ZX	9.189×10^{-1}	3.447×10^{4}
016003	2btTauW	5.472×10^{-1}	6.996×10^{4}
016004	2Photons	1.7×10^{-2}	2.9×10^{5}
016007	2X	1.092×10^{1}	7.292×10^{2}
Pythia + BlackMax ($M_P = 2.0$ TeV, c3), AMI tag: e660_s933_r1831			
017000	2Electrons	2.28×10^{-2}	2.04×10^{5}
017001	2Muons	2.25×10^{-2}	2.04×10^{5}
017002	1ZX	1.125×10^{0}	6.094×10^{4}
017003	2btTauW	7.550×10^{0}	4.512×10^{3}
017004	2Photons	1.1×10^{-2}	4.0×10^{5}
017005	ElectronMuon	4.50×10^{-2}	2.07×10^{5}
017006	1ElectronMuonX	1.472×10^{0}	3.663×10^{4}
017007	2X	3.100×10^{1}	8.495×10^{2}
Pythia + BlackMax ($M_P = 2.0$ TeV, c1), AMI tag: e660_s933_r1831			
018000	2Electrons	2.24×10^{-2}	4.47×10^{5}
018001	2Muons	2.26×10^{-2}	4.42×10^{5}
018002	1ZX	1.142×10^{0}	4.866×10^{4}
018003	2btTauW	7.940×10^{0}	4.819×10^{3}
018004	2Photons	1.1×10^{-2}	4.4×10^{5}
018005	ElectronMuon	4.51×10^{-2}	2.22×10^{5}
018006	1ElectronMuonX	2.259×10^{-1}	2.654×10^{5}
018007	2X	3.634×10^{1}	8.501×10^{2}

Table C.7: Release 16: PBHs

Appendix D

Transform options

The options given to each transform in producing the signal datasets are listed here.

D.1 Release 15 Atlfast-IID

Evgen_trf.py, Release 15.6.10.6

- EcmEnergy = 7000 csc_atlasG4_trf.py, Release 15.6.10.6

- geometryVersion = ATLAS-GEO-10-00-00
- physicsList = QGSP_BERT
- jobConfig = VertexFromCondDB.py, FastSimulationJobTransforms/jobConfig.onlyFastIDKiller.py
- conditionsTag = OFLCOND-DR-BS7T-ANom-15
- DBRelease = DBRelease-11.3.1.tar.gz

csc_digi_trf.py, Release 15.6.10.6

- geometryVersion = ATLAS-GEO-10-00-00
- minbiasHitsFile = NONE
- cavernHitsFile = NONE
- beamHaloHitsFile = NONE
- beamGasHitsFile = NONE
- jobConfig = FastSimulationJobTransforms/jobConfig.AtlfastIID.py
- digiRndmSvc = AtRanluxGenSvc
- samplingFractionDbTag = QGSP_BERT
- conditionsTag = OFLCOND-DR-BS7T-ANom-15
- DBRelease = DBRelease-11.3.1.tar.gz

Reco_trf.py, Release 15.6.10.6

- autoConfiguration = everything
- postExec = 'topSequence.TimeOut = 7200*Units.s'
- preExec = 'rec.Commissioning.set_Value_and_Lock(True); InDetFlags.doMinBias.set_Value_and_Lock(False); jobproperties.Beam.energy.set_Value_and_Lock(3500*Units.GeV); muonRecFlags.writeSDOs = True;rec.doDPD = True; InDetFlags.useBeamConstraint.set_Value_and_Lock(True);

from CaloRec.CaloCellFlags import jobproperties;
jobproperties.CaloCellFlags.doLArDeadOTXCorr.set_Value_and_Lock(True)'
- preInclude = RecJobTransforms/SetJetConstants-02-000.py,
RecJobTransforms/NoTrackSlimming.py
- conditionsTag = OFLCOND-DR-BS7T-ANom-15
- geometryVersion = ATLAS-GEO-10-00-00
- triggerConfig = MCRECO:DB:TRIGGERDBMC:145,73,124
- DBRelease = DBRelease-11.3.1.tar.gz

D.2 Release 16 full simulation

Evgen_trf.py, Release 15.6.13.11

- EcmEnergy = 7000
- EvgenJobOpts = MC10JobOpts-00-00-86.tar.gz

csc_atlasG4_trf.py, Release 15.6.12.9

- geometryVersion = ATLAS-GEO-16-00-00
- physicsList = QGSP_BERT
- jobConfig = VertexFromCondDB.py,
jobConfig.LooperKiller.py,CalHits.py
- conditionsTag = OFLCOND-SDR-BS7T-02
- DBRelease = DBRelease-12.2.1.tar.gz

Digi_trf.py, Release 16.0.2.6

- geometryVersion = ATLAS-GEO-16-00-00
- NDMinbiasHitsFile = mc10_7TeV.105001.pythia_minbias.merge.HITS.e577_s932_s952
- cavernHitsFile = mc10_7TeV.005008.CavernInput.merge.HITS.e4_e607_s951_s952
- beamHaloHitsFile = NONE
- beamGasHitsFile = NONE
- postExec = 'ToolSvc.LArAutoCorrTotalToolDefault.NMinBias=0'
- preInclude = 'SimuJobTransforms/PileUpBunchTrains2010Config8_DigitConfig.py'
- samplingFractionDbTag = QGSP_BERT
- conditionsTag = OFLCOND-SDR-BS7T-04-02
- DBRelease = DBRelease-13.1.1.tar.gz

Reco_trf.py, Release 16.0.2.6

- autoConfiguration = everything
- extraParameter = 'postExec_r2e'
- postExec_r2e = 'topSequence.TrigT1RPC.PatchForRpcTime=True'
- preExec = 'rec.Commissioning.set_Value_and_Lock(True);
jobproperties.Beam.energy.set_Value_and_Lock(3500*Units.GeV);
muonRecFlags.writeSDOs=True;
from CaloRec.CaloCellFlags import jobproperties;
jobproperties.CaloCellFlags.doMinBiasAverage=False;
jobproperties.Beam.numberOfCollisions.set_Value_and_Lock(2.2);
jobproperties.Beam.bunchSpacing.set_Value_and_Lock(150)'

- preInclude = 'RecJobTransforms/SetJetConstants-02-000.py'
- conditionsTag = OFLCOND-SDR-BS7T-04-02
- geometryVersion = ATLAS-GEO-16-00-00
- triggerConfig = MCRECO:DB:TRIGGERDBMC:248,108,194
- DBRelease = DBRelease-13.1.1.tar.gz

Appendix E

ADC plotting service

During summer 2010 I attended the Summer Student Programme at CERN. Besides following the lectures, I was put to the task of developing a proof of concept for a plotting service for the ATLAS Distributed Computing (ADC) monitoring. At that time, there where almost no consistency between the monitoring pages. The idea was to provide all pages with the desired graphs from one single server, easing maintenance. With excellent guidance from my supervisor on the project, Graeme Stewart, we ended up with a working prototype of a plotting service.

E.1 Tools

To carry out the project, we needed a web application framework and a plotting package. The choice fell on the Python-based Django [96] and matplotlib [97], and finally GraphTool [98], a wrapper for matplotlib. Python's clear, readable syntax, exception-based error handling, good introspection capabilities and neat standard libraries, makes it suitable for basing a service on. I will shortly describe the mentioned packages, their functionality and my work with them in the following.

E.1.1 Web framework

Django is a high-level web framework designed to make common web-development fast and easy. Getting started is simply done by typing

```
django-admin.py startproject mysite
```

from the command line. A directory named *mysite* is created, containing scripts for management, configuration and URL declarations for the project. Doing

```
python manage.py startapp myapp
```

will further create a subdirectory named *myapp*, containing the *views.py* to be filled with callback functions. The *models.py* is suitable for declaring classes of graphs. While Django has many elegant features, I mainly worked with the URL dispatcher and the way Django handles http requests.

The URL configuration module is a mapping between URL patterns, simple regular expressions, to Python callback functions, which are defined in the views. There are no restrictions on the length of the mapping, it can refer to other mappings and be constructed dynamically. When a user requests a page, Django searches for the first pattern matching the requested URL and calls its respective view. As a first argument, an instance of the HttpRequest class is passed to the views. Within this object there is a dictionary-like object, a QueryDict, containing all

given HTTP GET or POST parameters. The parameters, e.g. title, x-label, font or title size of a plot, can be specified in the URL with the following syntax

```
key1=value1&key2=value2&key3=value3&...
```

As a temporary solution for passing the data to be plotted, we used the JSON package included in the Python standard library from Python v2.6. The data had to be written as a (nested) Python dictionary and sent as a JSON encoded string,

```
mydata_string = json.dumps(mydata_dict).
```

It was then possible to send the data in the URL by using the same syntax as mentioned earlier,

```
data=mydata_string&...
```

The views parse the QueryDict by JSON encoding the string associated with the data key and converting parameters expected to be numbers to floats. All the information is then passed to the graph object to be created. The final graph is stored as a temporary png file and returned to Django as an HttpResponse object.

E.1.2 Plotting tools

The matplotlib package was developed with a wish of emulating the MATLAB graphics commands. As Python is not known for being the fastest of programming languages, matplotlib makes heavy use of NumPy and other extensions giving good performance even for large arrays of data. matplotlib provides a wide arrange of publication-quality 2D graphs that can be used in web application servers.

GraphTool, the wrapper for matplotlib, was developed by CMS for generating high-quality graphs with only a few lines of code. It contains the graph types most frequently used by the ATLAS monitoring pages; the stacked bar graph, the cumulative graph and the pie chart. In addition to these, the quality map was also desired.

I gave each graph type an explicit URL with a corresponding view function. After parsing the QueryDict, creating a GraphTool graph isn't much of a hassle. Just do

```
data, metadata = parse(HttpRequest) mygraph = GraphType()
mygraph(data, filename, metadata),
```

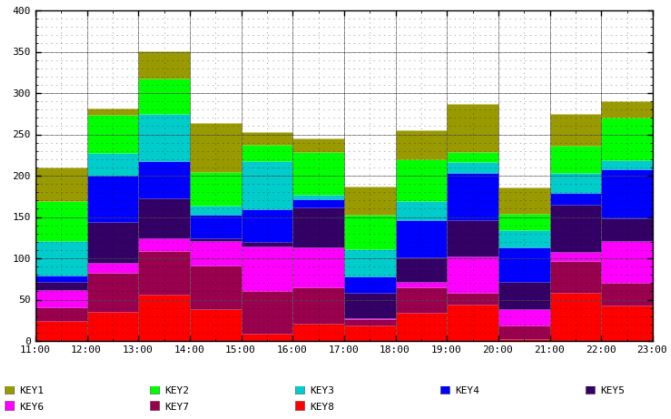
and the graph is stored in the file specified by the filename variable.

In models.py I defined subclasses of StackedBarGraph (SBG), QualityMap (QM), PieGraph (PG) and CumulativeGraph (CG) and made some changes mainly to the appearance of the graphs. The default colours were hard-coded in SBG, PG and CG, and thus not very flexible. ATLAS now has its own colour scheme to be used, but at that time I found it useful to have a function generating as many colours as needed. QM originally used a colour map ranging from red to green, which is unfortunate for colour-blind persons. Therefore I changed the colour map to one ranging from white to very dark purplish with nice shades of orange in between. In the SBG I made sure that the keys, being ATLAS clouds, timestamps etc., are always sorted alphabetically or numerically. Both QM and CG were already subclasses of TimeGraph (TG), which is a special class for graphs with times or dates on their x-axis. TG utilizes the clever functions PrettyDateLocator and PrettyDateFormatter from matplotlib, displaying the timestamp info in a most sensible way. I defined the class TimeStackedBarGraph inheriting from TG and SBG, making also SBG suitable for plotting time series data.

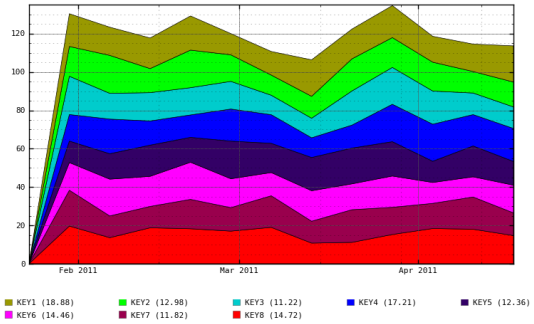
E.2 The result

In Figure E.1 an example of each graph is shown for randomly generated series of data. The Python-based plotting service was at the end set aside in favour of Flot [99], a pure Javascript

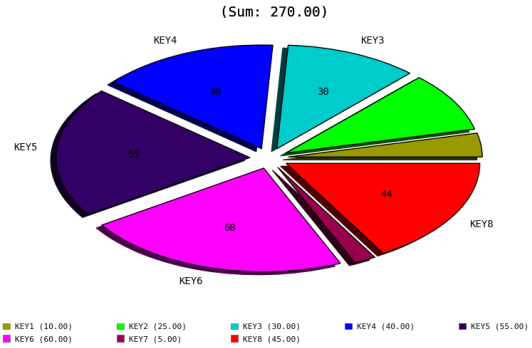
plotting library for jQuery. The key features of Flot are simple usage, attractive and interactive, where the latter allows for e.g. zooming and mouse tracking. The Flot library has been successfully implemented in the PanDA (Production and Distributed Analysis System) monitoring page.



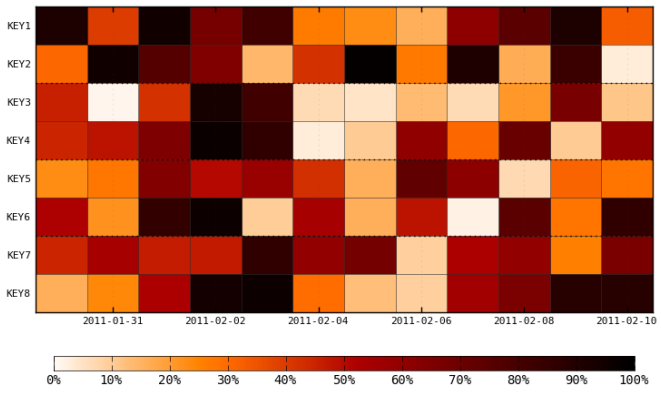
(a) Stacked bar graph example



(b) Cumulative graph example



(c) Pie Graph example



(d) Quality map example

Figure E.1: Example plots for randomly generated data.

Bibliography

[1] P.A.M. Dirac. The Quantum Theory of the Emission and Absorption of Radiation. *Proceedings of the Royal Society of London, Series A*, 114, 1927.

[2] W.N Cottingham, D.A. Greenwood. *An Introduction to the Standard Model of Particle Physics*. Cambridge, 2007.

[3] F. Mandl, G. Shaw. *Quantum Field Theory (Revised Edition)*. Wiley, 1993.

[4] F. Englert, R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Physical Review Letters*, 13, 1964.

[5] P. Higgs. Broken symmetries, massless particlees and gauge fields. *Physics Letters*, 12, 1964.

[6] P. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Physical Review Letters*, 13, 1964.

[7] G. S. Guralnik, C. R. Hagen, T. W. B. Kibble. Global Conservation Laws and Massless Particles. *Physical Review Letters*, 13, 1964.

[8] G. 't Hooft. Renormalizable Lagrangians for Massive Yang-Mills Fields. *Nuclear Physics B*, 35, 1971.

[9] F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6, 1933.

[10] N. Jarosik et al. Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Sky Maps, Systematic Errors, and Basic Results. *The Astrophysical Journal Supplement Series*, 192, 2011.

[11] N. Arkani-Hamed, S. Dimopoulos, G. Dvali. The Hierarchy Problem and New Dimensions at a Millimeter. *Physics Letters B*, 429, 1998.

[12] L. Randall, R. Sundrum. A Large Mass Hierarchy from a Small Extra Dimension. *Physical Review Letters*, 83, 1999.

[13] S. B. Giddings, S. Thomas. High Energy Colliders as Black Hole Factories: The End of Short Distance Physics. *Physical Review D*, 65, 2002.

[14] G. F. Giudice, R. Rattazzi, J. D. Wells. Quantum Gravity and Extra Dimensions at High-Energy Colliders. *Nuclear Physics B*, 54, 1999.

[15] S. Dimopoulos, G. Landsberg. Black Holes at the LHC. *Physical Review Letters*, 87, 2001.

[16] D. J. Kapner et al. Tests of the Gravitational Inverse-Square Law below the Dark-Energy Length Scale. *Physical Review Letters*, 98, 2007.

[17] A. Stefan. Extra Dimension Search at LEP. In *32nd International Conference on High Energy Physics*, 2004.

[18] Edgar Carrera for the D0 Collaboration. Physics with Single Photons plus Missing Energy Final States at D0. In *Lake Louise Winter Institute: Fundamental Interactions (LLWI 2009)*, 2009.

[19] The CDF Collaboration, T Aaltonen et al. Search for Large Extra Dimensions in Final States Containing One Photon or Jet and Large Missing Transverse Energy Produced in pp Collisions at $\sqrt{s} = 1.96$ TeV. *Physical Review Letters*, 101, 2008.

[20] P. Kanti. Black Holes in Theories with Large Extra Dimensions: a Review. *International Journal of Modern Physics A*, 19, 2004.

[21] G. Landsberg. Black Holes at Future Colliders and Beyond: a Review. *Journal of Physics G*, 32, 2006.

[22] M. Cavaglià. Black Hole and Brane Production in TeV Gravity: a Review. *International Journal of Modern Physics A*, 18, 2003.

[23] S. Hossenfelder. What Black Holes Can Teach us. In *Focus on Black Hole Research*. Nova Science Publishers, 2005.

[24] K. S. Thorne. Nonspherical gravitational collapse: a short review. In *Magic without Magic*. Freeman, San Francisco, 1972.

[25] H. Yoshino, Y. Nambu. Black hole formation in the grazing collision of high-energy particles. *Physical Review D*, 67, 2003.

[26] H. Yoshino, V. S. Rychkov. Improved analysis of black hole formation in high-energy particle collisions. *Physical Review D*, 71, 2005.

[27] R. C. Myers, M. J. Perry. Black Holes in Higher Dimensional Space-Times. *Annals of Physics*, 172, 1986.

[28] P. Meade, L. Randall. Black holes and quantum gravity at LHC. *Journal of High Energy Physics*, 53, 2008.

[29] P. D. D'Eath. Gravitational radiation in high-speed black-hole collisions. *Classical and Quantum Gravity*, 10, 1993.

[30] R. Emparan, G. T. Horowitz, R. C. Myers. Black Holes Radiate Mainly on the Brane. *Physical Review Letters*, 85, 2000.

[31] S. Hossenfelder, M. Bleicher, S. Hofmann, H. Stöcker, A. V. Kotwal. Black Hole Relics in Large Extra Dimensions. *Physics Letters B*, 566, 2003.

[32] B. Koch, M. Bleicher, S. Hossenfelder. Black Hole Remnants at the LHC. *Journal of High Energy Physics*, 53, 2005.

[33] H. Stöcker. Stable TeV - Black Hole Remnants at the LHC: Discovery through Di-Jet Suppression, Mono-Jet Emission and a Supersonic Boom in the Quark-Gluon Plasma. *International Journal of Modern Physics D*, 16, 2007.

[34] D. Stojkovic, G. D. Starkman, F. C. Adams. Information-preserving black holes still do not preserve baryon number and other effective global quantum numbers. *International Journal of Modern Physics D*, 14, 2005.

[35] The ATLAS Collaboration, G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3, 2008.

[36] The ATLAS Collaboration. *ATLAS Computing: Techical Design Report.* CERN, 2005.

[37] Dai et al. BlackMax: A black-hole event generator with rotation, recoil, split branes, and brane tension. *Physical Review D*, 77, 2008.

[38] T. Sjöstrand, S. Mrenna, P. Skands. PYTHIA 6.4 physics and manual. *Journal of High Energy Physics*, 5, 2006.

[39] ATLAS Computing. Athena Framework. `https://twiki.cern.ch/twiki/bin/view/Atlas/AthenaFramework`.

[40] C. Ay et al. Monte Carlo generators in ATLAS software. *Journal of Physics: Conference Series*, 219, 2010.

[41] M. Dobbs, J. B. Hansen. The HepMC C++ Monte Carlo event record for High Energy Physics. *Computer Physics Communications*, 134, 2001.

[42] The GEANT4 Collaboration, S. Agostinelli et al. GENAT 4 - a simulation toolkit. *Nuclear Instruments and Methods in Physics Research A*, 506, 2003.

[43] D. Adams et al. The ATLFAST-II performance in release 14 -particle signatures and selected benchmark processes-. Technical Report ATL-PHYS-INT-2009-110, CERN, Geneva, December 2009.

[44] M. Dührssen. The fast calorimeter simulation FastCaloSim. Technical Report ATL-PHYS-INT-2008-043, ATL-COM-PHYS-2008-093, CERN, Geneva, July 2008.

[45] The ATLAS Collaboration, G. Aad et al. *Expected performance of the ATLAS experiment - detector, trigger and physics.* CERN, 2008.

[46] Rene Brun, Fond Rademarkers. ROOT - An Object Oriented Data Analysis Framework. *Nuclear Instruments and Methods in Physics Research A*, 389, 1997.

[47] P. M. Nadolsky et al. Implications of CTEQ global analysis for collider observables. *Physical Review D*, 78, 2008.

[48] Dirk Duellmann[1] The LCG POOL Project, General Overview and Project Structure. In *Computing in High Energy and Nuclear Physics (CHEP03)*, 2003.

[49] ATLAS Computing. Top level tags in the ATLAS Geometry Database. `https://twiki.cern.ch/twiki/bin/viewauth/Atlas/AtlasGeomDBTags`.

[50] ATLAS Computing. Cool Prod Tags. https://twiki.cern.ch/twiki/bin/view/Atlas/CoolProdTags.

[51] N. Arkani-Hamed, M. Schmaltz. Hierarchies without symmetries from extra dimensions. *Physical Review D*, 61, 2000.

[1] What a great name for a Nowegian wrestling character!

[52] Dai et al. *Manual of BlackMax, a black-hole event generator with rotation, recoil, split branes, and brane tension*, Version 2.00 edition, 2009.

[53] M. Whalley, A. Buckley. LHAPDF the Les Houches Accord PDF Interface. `http://projects.hepforge.org/lhapdf/`.

[54] ATLAS Computing. Atlfast II. `https://twiki.cern.ch/twiki/bin/viewauth/Atlas/AtlfastII`.

[55] The ATLAS Collaboration, M. Beckingham et al. The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim. Technical Report ATL-PHYS-PUB-2010-013, CERN, Geneva, October 2010.

[56] W. Lampl et al. Calorimeter clustering algorithms: Description and performance. Technical Report ATL-LARG-PUB-2008-002, ATL-COM-LARG-2008-003, CERN, Geneva, April 2008.

[57] M. L. Mangano et al. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *Journal of High Energy Physics*, 7, 2003.

[58] J. Butler, T. Nunnemann, J. Qian. NNLO Drell-Yan Cross Sections for Background Normalization. `http://www-personal.umich.edu/~qianj/DrellYanStudySlides.pdf`.

[59] J. Alwall et al. MadGraph/MadEvent v4: The New Web Generation. *Journal of High Energy Physics*, 9, 2007.

[60] S. Frixione, B.R. Webber. Matching NLO QCD computations and parton shower simulations. *Journal of High Energy Physics*, 6, 2002.

[61] S. Frixione, P. Nason, B.R. Webber. Matching NLO QCD and parton showers in heavy flavour production. *Journal of High Energy Physics*, 8, 2003.

[62] G. Corcella et al. HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes). *Journal of High Energy Physics*, 1, 2001.

[63] The ATLAS Collaboration. Search for New Physics in Multi-Body Final States at High Invariant Masses with ATLAS. Technical Report ATLAS-CONF-2010-088, CERN, Geneva, October 2010.

[64] W,Z Physics Working Group. ATLAS WZ physics working group. `https://twiki.cern.ch/twiki/bin/view/AtlasProtected/SMWZSignatures`.

[65] The ATLAS Collaboration, G. Aad et al. Readiness of the ATLAS Liquid Argon Calorimeter for LHC Collisions. *The European Physical Journal C*, 70, 2010.

[66] The ATLAS Collaboration, G. Aad et al. Search for a heavy particle decaying into an electron and a muon with the ATLAS detector in $\sqrt{s} = 7$ TeV pp collisions at the LHC. *Physical Review Letters*, 2011.

[67] A. L. Read. Presentation of search results: the CL_s teqhinque. *Journal of Physics G*, 28, 2002.

[68] G. Cowan, K. Cranmer, E. Gross, O. Vitells. Power-Constrained Limits. arXiv:1105.3166, May 2011.

[69] V. Bartsch, G. Quast. Expected signal observability at future experiments. Technical Report CMS-NOTE-2005-004, CERN-CMS-NOTE-2005-004, CERN, Geneva, February 2005.

[70] J. Neyman, E. S. Pearson. On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231, 1933.

[71] A. L. Read. Optimal Statistical Analysis of Search Results based on the Likelihood Ratio and its Application to the Search for the MSM Higgs Boson at $\sqrt{s}$ = 161 and 172 GeV. *DELPHI 97-158 PHYS 737*, 1997.

[72] G. Cowan, K. Cranmer, E. Gross, O. Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal*, 71, 2011.

[73] R. Blair et al. ATLAS Standard Model Cross Section recommendations for 7 TeV LHC running. `https://svnweb.cern.ch/trac/atlasgrp/browser/Physics/StandardModel/xsectf/note/xsectf.pdf`.

[74] R. Hamberg, W. L. van Neerven, T. Matsuura. A complete calculation of the order α_s^2 correction to the Drell-Yan K-factor. *Nuclear Physics B*, 359, 1991.

[75] J. Butterworth et al. Single Boson and Diboson Production Cross Sections in pp Collisions at $\sqrt{s}$ = 7 TeV. Technical Report ATL-COM-PHYS-2010-695, CERN, Geneva, August 2010.

[76] C. Anastasiou, L. Dixon, K. Melnikov, F. Petriello. High-precision QCD at hadron colliders: electroweak gauge boson rapidity distributions at NNLO. *Physical Review D*, 69, 2004.

[77] J. M. Campbell, R. K. Ellis. An update on vector boson pair production at hadron colliders. *Physical Review D*, 60, 1999.

[78] The ATLAS Collaboration, G. Aad et al. Measurement of the top quark-pair production cross section with ATLAS in pp collisions at $\sqrt{s}$ = 7 TeV. Technical Report CERN-PH-EP-2010-064, CERN, Geneva, December 2010.

[79] A. A. Abdelalim et al. Search for high mass dielectron resonances at $\sqrt{s}$ = 7 TeV. Technical Report ATL-COM-PHYS-2011-083, CERN, Geneva, February 2011.

[80] M. Baak et al. Data Quality Status Flags and Good Run Lists for Physics Analysis in ATLAS. Technical report, CERN, Geneva, February 2009.

[81] The ATLAS collaboration. Updated Luminosity Determination in pp Collisions at $\sqrt{s}$ = 7 TeV using the ATLAS Detector. Technical Report ATLAS-CONF-2011-011, CERN, Geneva, March 2011.

[82] The ATLAS Collaboration, G. Aad et al. Luminosity Determination in *pp* Collisions at $\sqrt{s}$ = 7 TeV using the ATLAS Detector at the LHC. Technical Report CERN-PH-EP-2010-069, CERN, Geneva, January 2011.

[83] EGamma group. Electrons Energy Direction. `https://twiki.cern.ch/twiki/bin/view/AtlasProtected/ElectronsEnergyDirection`.

[84] EGamma group. IsEM Identification. `https://twiki.cern.ch/twiki/bin/view/AtlasProtected/IsEMIdentification`.

[85] EGamma group. Efficiency Measurements. https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EfficiencyMeasurements.

[86] EGamma group. Energy Rescaler. https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EnergyRescaler.

[87] EGamma group. Energy Scale Resolution Recommendations. https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EnergyScaleResolutionRecommendations.

[88] Lepton +X Exotics Working Group. Cutflows Zprime Rel16. https://twiki.cern.ch/twiki/bin/view/AtlasProtected/CutflowsZprimeRel16.

[89] Muon Combined Performance. Guidelines for Analysis in Release 16 . https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MCPAnalysisGuidelinesRel16.

[90] F. Cerutti et al. Muon Momentum Resolution in First Pass Reconstruction of pp Collision Data Recorded by ATLAS in 2010. Technical Report ATLAS-COM-CONF-2011-003, CERN, Geneva, January 2011.

[91] Lepton +X Exotics Working Group. Cutflows Zprime MuMu Rel16. https://twiki.cern.ch/twiki/bin/view/AtlasProtected/CutflowsZprimeMuMuRel16.

[92] W. Verkerke, D. Kirkby. The RooFit toolkit for data modeling. In *PHYSTAT 05: Statistical Problems in Particle Physics, Astrophysics and Cosmology*, 2005.

[93] F. James. MINUIT, Function Minimization and Error Analysis, Reference Manual Version 94.1. In *CERN Program Library Long Writeup D506*. CERN, 1994.

[94] Roger Barlow. Extended maximum likelihood. *Nuclear Instruments and Methods in Physics Research A*, 297, 1990.

[95] J. S. Conway. Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra. Technical Report arXiv:1103.0354, CERN, March 2011. Presented at PHYSTAT 2011, CERN, Geneva, Switzerland, January 2011, to be published in a CERN Yellow Report.

[96] Django. http://www.djangoproject.com/.

[97] matplotlib. http://matplotlib.sourceforge.net/.

[98] Graphtool. http://t2.unl.edu/documentation/graphtool.

[99] Flot. http://code.google.com/p/flot/.